

Drugs are for beauty too

Like it or not, many people want to improve their looks with the help of biology, and pharmaceutical companies look set to help them. Regulatory bodies need to catch up.

They want to be pert, they want to be taut, they want glossy manes into their 60s. And they want balms, dyes and depilatories that will help them achieve it. Some might dismiss it as vanity, but society's increasing preoccupation with looks is fuelling a booming business in cosmetic drugs, or 'cosmeceuticals', worth \$3.4 billion last year in the United States alone.

Cosmeceuticals lurk in the shadowy ground between drugs and cosmetics. Allergan's Botox, which flattens furrowed foreheads, is one example. Merck's Propecia for balding pates is another, as are off-the-shelf skin creams with active biological ingredients.

Those who add a prescription cosmeceutical to their morning routine can have real hope of seeing results. But the enticing skincare aisles of your local drugstore tell a different story: dermatologists confess that some 90% of ingredients in anti-ageing creams are little more than overpriced petroleum jelly. Why, when you slather on jelly for wrinkles, do you need to smear on a healthy dose of scepticism?

Cosmeceutical manufacturers must shoulder much of the blame, for trying to sidestep drug regulations. The US Food and Drug Administration (FDA) defines drugs as agents intended for use in the diagnosis, cure, mitigation, treatment or prevention of disease, or that affect the structure or function of the body. Cosmetics escape the rigorous trials demanded of a drug because they are assumed only to alter our appearance. But the system falls down for cosmeceuticals, because it is the manufacturers — not the FDA — who decide whether a product is classed as drug or cosmetic. A cream that claims to cure eczema is a drug; if it claims to promote healthy skin, it is a cosmetic. This bizarre situation means that, although some cosmetics companies have excellent R&D arms, many are dissuaded from

finding genuinely active ingredients, or advertising their properties if they do, for fear of having to undergo expensive drug trials. This also fuels the pseudo-science that is used to hype cosmetics.

This state of affairs must change. Cosmeceutical manufacturers should show that their ingredients genuinely work, or find some that do. The FDA should revise its outdated regulations to enforce this, so that all biologically active ingredients, whether drugs or cosmetics, have to prove their worth. Consumers are helpless to tell fact from fiction in today's non-prescription creams, so there is a market for prescription drugs with proven cosmetic powers. A few pharmaceutical and cosmetics giants, including Pfizer and L'Oreal, are embarking on research to fill this niche. Some people might argue that pharmaceutical companies should concentrate on curing deadly diseases, not feeding society's body paranoia. But if there is money in the beauty parlour, they are likely to pursue it (see page 990).

Perhaps a more thorny dilemma will face doctors who have to decide whether to prescribe cosmetic drugs to patients who demand them for wrinkles or baldness. As with any non-essential therapy, national or private health insurers can exclude them from coverage. But doctors and medical ethicists must consider whether they have the right to refuse a prescription to patients willing to pay.

The dilemmas thrown up by cosmeceuticals will become more pressing as biologists reveal the secrets behind wilting skin and barren follicles — and the molecules that might revamp them. Our attitudes are outdated. Consumers should demand more from their \$20-a-pot unguents. And when 'real' cosmeceuticals are created, doctors, industry and regulators should work together to ensure that they are effective and accessible. ■

Fetal trials need public funds

The outcomes of trials of fetal-cell transplants highlight the importance of public access.

Last week, neurologists finally published the long-awaited results of a clinical trial that has been controversial since its outset (see page 987). The study was the second of two funded by the US National Institutes of Health (NIH) to look at whether fetal-cell transplants could ease the symptoms of Parkinson's disease. The controversy continues, as neuroscientists argue over what the results mean, but the bottom line is that public funding in this emerging field ensured that the trials were done in the most open, rigorous, ethical and safe way possible. Policy-makers and funding agencies should keep this lesson in mind, especially in a political climate that is increasingly hostile towards other areas of research, such as work on human embryonic stem cells.

The Parkinson's trials were first opposed because they used tissue from aborted fetuses. The US government could not fund such trials because of restrictions preventing scientists from using federal funding for work on fetal tissue. But after then President Bill Clinton lifted the restriction on federal funding for fetal-tissue research, scientists began two NIH-funded trials of the technique.

In 2001, a publication on the first Parkinson's trial showed that

some patients had developed involuntary muscle movements called 'dyskinesias'. The second trial found the same side-effects. But these were publicly funded trials, and the researchers were free to discuss their results in journals and at meetings, so scientists now have some idea about what caused the dyskinesias. More importantly, they have publicly available data to help them evaluate their ideas.

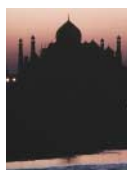
This is a very different situation from some privately funded trials, such as an Alzheimer's-vaccine trial that was stopped last year when some patients developed life-threatening brain inflammation in a study sponsored by a company called Elan. The side-effect was reported in a company press release, not at a scientific meeting. The scientific community, short of data, is still trying to work out what went wrong.

This is a crucial lesson for countries such as the United States that are trying to decide how much to get involved in work on human embryonic stem cells. Whether or not governments fund such work, investors will eventually fund private trials themselves. If policy-makers want to ensure that the data from such trials do not disappear into a black hole, and that trials are done as safely as possible, it should err on the side of more, not less, funding for the work. ■





Closed doors
Big business quashes talks on 'open' innovation
p984



Aiming high
India's plans for Moon mission cleared for take-off
p985



Safety catch
Homeland Security leaves researchers strapped for cash
p986



Fetal flaw
Tissue transplants for Parkinson's fail to impress
p987

Quest for SARS source gathers pace in bid to thwart resurgence

Alison Abbott, Munich

At least three species of wild animals sold in markets in the Guangdong province of southern China have been shown to harbour the virus that causes severe acute respiratory syndrome (SARS). And a substantial number of similar leads are being developed, according to an international delegation of experts.

The delegation was in Guangdong, where the SARS outbreak began, this month to assess the progress being made in the hunt for the disease's animal 'reservoir'. Like flu, SARS is suspected to be seasonal. But the group found that as winter approaches, local scientists are a long way from identifying which species is the true reservoir for the disease — and so could spark a fresh SARS outbreak.

Several Chinese research institutes have been involved in the intense effort to test both domesticated and wild animals for the coronavirus that causes SARS. And the delegation, split half-and-half between Chinese researchers and those from elsewhere, representing the World Health Organization (WHO) and the United Nations Food and Agriculture Organization, was surprised by some of the results it was shown.

At a press conference on 20 August, the visiting scientists discussed the wide range of animals — from snakes through to mammals — that have been tested for the SARS coronavirus so far. Most of the results were negative, but a surprising number of species tested positive, the group said.

The animals were subjected to different tests based on the polymerase chain reaction (PCR), which amplifies small stretches of genetic material from the SARS coronavirus and allows them to be detected. Some were also tested for the presence of antibodies to the human SARS virus.

The relatively high number of positive results is unusual, the experts said. Normally just one main and a few secondary reservoirs would be expected. "These results call into question the specificity of the tests," says François Moutou, head of the French food-safety agency's epidemiological unit and a member of the group. "It would be unusual if species as diverse as mammals, birds and reptiles were all harbouring viruses closely



Animals on sale in Guangdong's markets are offering leads to possible sources of the SARS virus.

related to the human SARS virus." Real confirmation of the presence of a SARS-related virus, and a determination of its relationship to the human SARS virus, requires isolation and sequencing of the entire viral genome, he points out.

Some groups have already done this. For example, in work to be published shortly, scientists in Hong Kong have completed DNA sequences of coronaviruses from several different species that turn out to be closely related to the human SARS virus. The WHO has recommended that this work be extended as a priority. The results so far do not indicate whether animals infected humans or vice versa, or how the virus may have jumped between species. When more such sequences are completed, an evolutionary tree for the virus could be constructed that could indicate where it first began to evolve, thus helping to identify the reservoir. Once identified, measures could be put in place to isolate it from human contact.

But even in the absence of complete information about the reservoir, the expert delegation recommended that regulations on the

farming, trading and consumption of wild animals in China should be strengthened.

Meanwhile, alarm bells have been switched off in Canada, where the virus responsible for an outbreak last month of a mild respiratory illness at a nursing home in Surrey, British Columbia, has been found not to be closely related to the SARS virus.

Although symptoms of the illness had not been typical of SARS, many patient specimens had tested positive in SARS tests based on the PCR reaction. Health workers were concerned that the SARS virus could have mutated to a less virulent form, which would have implied that it mutates more rapidly than had been thought, and so could mutate back into a virulent form.

The PCR tests rely on 'primers' that correspond to short sequences of the SARS virus. The false positives occurred because the Canadian virus is a member of the same family as the SARS virus, and so shares some of its DNA. But once researchers in British Columbia had sequenced more of the virus they found differences that showed that the two were only distant cousins. ■

Business backlash kills off software meeting

Declan Butler

Lobbying by US business interests has shot down a proposed meeting, to be held by the Geneva-based World Intellectual Property Organization (WIPO) next year, to explore 'open' models of innovation.

A group of leading scientists and economists suggested the meeting in a 7 July letter to Kamil Idris, director-general of WIPO. They highlighted the explosion of various open and collaborative projects to create public goods, such as the Human Genome Project and open-source software, and said it was time for WIPO to include the approach in its deliberations.

Francis Gurry, assistant director-general and legal counsel at WIPO, told *Nature* at the time that the agency welcomed the idea as "a very important and interesting development". He added that "the director-general looks forward with enthusiasm to taking up the invitation to organize a conference to explore the scope and application of these models" (see *Nature* **424**, 118; 2003).

But Gurry's words seem to have triggered a backlash from firms that would rather see WIPO working to protect their intellectual property rights. Gurry says WIPO has since been inundated with calls from trade and consumer groups and government representatives. It is understood that lobbyists such as the Business Software Alliance, which is partly funded by Microsoft, also pressed the US state department and the US Patent and Trademark Office to have the meeting called off. US government officials have since spoken out against the idea.

"The request for an open discussion on a range of projects became transformed into a domestically, as opposed to internationally, polarized debate about just one 'project': open-source software," says Gurry. As a result, he says, the meeting is now unlikely to take place.



Empty seats: lobbying means the Linux penguin will not witness a discussion of intellectual property.

"The possibility of conducting a policy discussion on intellectual property of the sort appropriate for an international organization became increasingly remote," he claims.

In the United States, the European Union and elsewhere, governments are considering measures to encourage the public procurement of open-source software, such as the Linux operating system. This is being vigorously opposed by some software companies and by the Business Software Alliance.

Economist Paul David of Stanford University in California, who signed the letter, says he is "appalled" that the meeting should be "scuttled because of the mere presence of open-source software in a list of many other forms of collaborative knowledge production".

But Lois Boland, director of international relations for the US Patent and Trademark

Office, says that open-source software is contrary to WIPO's mission to promote intellectual property rights. "To hold a meeting to disclaim or waive such rights seems to us to be contrary to the goals of WIPO," she says.

The Washington-based Computer and Communications Industry Association, which represents small companies that support open-source software, has attacked WIPO's decision not to hold the meeting. Edward Black, president of the group, says that the association has written to Boland to complain about her comments.

Tim Hubbard, a genomicist at the Sanger Institute near Cambridge, UK, and a signatory to the letter, says he is still hopeful that the meeting may happen. "It's not the role of industries regulated by such regimes to inhibit discussion," he argues. ■

Biophysicist named to run basic-research arm of NIH

Erika Check, Washington

The US National Institute of General Medical Sciences (NIGMS)—the arm of the National Institutes of Health that focuses on basic research—has named its next director.

Jeremy Berg, head of biophysics and biophysical chemistry at Johns Hopkins University School of Medicine in Baltimore, Maryland, will take up the post in November. He succeeds Marvin Cassman, who left in May 2002 to head the Institute for Quantitative Biomedical Research, a biology lab at the University of California, San Francisco.



Jeremy Berg: varied approach.

Berg says he aims to find the right mix of investigator-initiated research and large-scale projects, including collaborative grants in areas such as cell signalling. "The spirit of investigator-initiated, curiosity-driven research has been what NIGMS is all about," he says. "But there are clearly advantages to large projects, so there has to be a balance."

Berg also professes a strong interest in training young scientists, and in the multi-institute projects planned by National Institutes of Health (NIH) director Elias

Zerhouni. He says that his previous work with Zerhouni, who was executive vice-dean at Johns Hopkins medical school before coming to the NIH, was an incentive for him to take the post at the NIGMS.

With an annual budget of more than \$1.8 billion, the NIGMS is one of the world's largest supporters of lab-based biological research. Unlike most branches of the NIH, it has no disease-specific remit but instead backs basic science in areas such as cell biology.

Berg's own research focuses on the structure and function of proteins, and in particular on how proteins recognize their target destinations. He plans to take his laboratory with him to the NIH campus in Bethesda, Maryland. ■

Ant book deepens divide over web publishing

Rex Dalton, San Diego

A disagreement about ants has highlighted increasing conflict between biologists and book publishers over the release of scientific monographs in print and online.

Brian Fisher, an entomologist at the California Academy of Sciences in San Francisco, is pressing for permission to publish data about ant species on the Internet. Under the terms of a book deal he signed last August with Harvard University Press, he cannot put material from his forthcoming monograph online for at least four years after it is printed.

The argument is just one example of the tension that is pervading several fields of systematic biology, researchers say. Many systematists want to publish their data and images on the web at the same time as they publish their monographs — hefty books that can document years of research. But publishers fear that simultaneous web publishing will reduce sales of the high-cost monographs.

“We are on the cusp of a renaissance in systematics,” says biologist Edward Wilson of Harvard University in Cambridge, Massachusetts. “But we are in a transition period of one form of publishing to another.”

Fisher’s deal with the Harvard press involves a monograph on the ants of Madagascar, where the isolated and diverse ecosystem is of special interest to systematic

biologists. Fisher hopes that the book will be published next year.

But Fisher also recently helped to launch AntWeb (www.antweb.org), a website that includes photographs of ants from Madagascar and California (see *Nature* **424**, 242; 2003). Harvard press officials are resisting his attempts to publish much of his data online before the monograph is published, worried that it will dent the book’s future sales.

“I don’t think the web release of material will hurt book sales; it will actually increase them,” Fisher says. Other researchers cite the US National Academies Press as an example of a publisher whose free online publication of its studies boosts its print sales. Correspondence shows that Harvard press disagrees, and is concerned about its ability to recover its costs in producing the monograph if AntWeb publishes much of its contents. Harvard press officials declined to comment on the dispute.

“These are difficult questions,” says Lynne Withey, director of the University of California Press. “People disagree about whether the web hurts or helps.” Officials at the California publisher are studying 30 of their social-science and humanities books to determine the impact of online publishing on a traditional monograph.

The recent publication by Harvard University Press of Wilson’s book *Pheidole in*



Edward Wilson believes that the age of the printed monograph could be at an end.

the New World: A Dominant, Hyperdiverse Ant Genus also brought criticism from some quarters about the lack of immediate free web access to its contents (see *Nature* **424**, 727; 2003). He says that the publisher is now putting the book online.

Wilson thinks that the best solution is for book publishers to put monographs online 6–12 months after print publication. He says that his latest book might be one of the last of its kind — “one of the last of the great sailing ships”, as he puts it, adding: “We need to work out some arrangement with publishers.” ■

India launches plan for mission to map the Moon

K. S. Jayaraman, New Delhi

India has announced that it will send an unmanned satellite to the Moon by 2008 — but observers are divided over the wisdom of the US\$100-million project.

The mission, called Chandrayan-1, was first mooted by the Indian Space Research Organisation (ISRO) in 1999. It will put a 525-kilogram satellite in polar orbit 100

kilometres above the Moon’s surface.

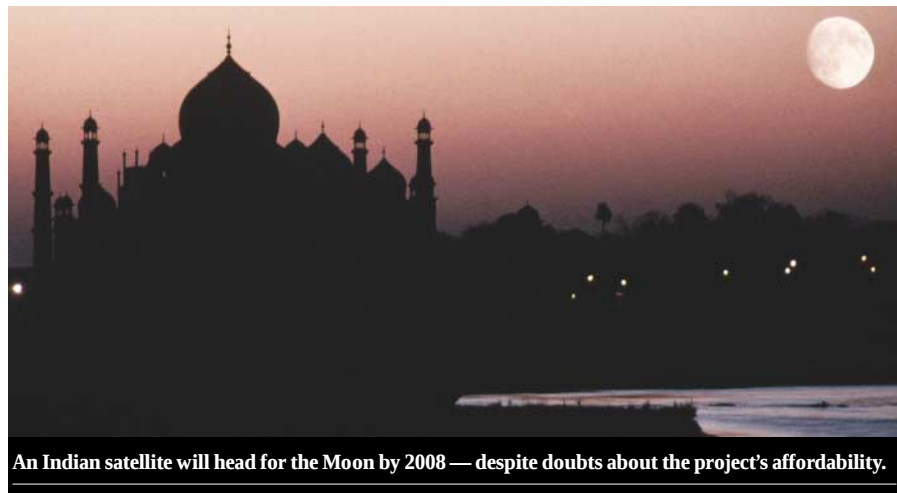
Prime Minister Atal Bihari Vajpayee gave the mission the go-ahead during his Independence Day address to the nation on 15 August. His speech was made as China prepares to steal a march on its Asian rival by sending a manned spacecraft into orbit this autumn.

The ISRO claims that the project has the

overwhelming support of India’s scientific community. Jayant Narlikar, former director of the Inter-University Centre for Astronomy and Astrophysics at Pune, says that it offers an important “intellectual challenge” to the country and will result in technological spin-offs that will benefit society.

But not everyone agrees. D. Raghuraman, a physicist at the non-profit Delhi Science Forum, says that the mission is a “luxury India can hardly afford”. And Roddam Narasimha, director of the National Institute of Advanced Studies in Bangalore, says that India would be better off doing lunar exploration as part of an international collaboration. He thinks that India should concentrate its attention on what is happening on Earth.

Chandrayan-1 will map the chemical composition of the entire lunar surface, and produce a three-dimensional atlas of regions of interest. In a statement, the ISRO said that the mission is a forerunner for other planetary missions, including “landing robots on the Moon and visits by Indian spacecraft to other planets in the Solar System”. ■



An Indian satellite will head for the Moon by 2008 — despite doubts about the project’s affordability.

B. KRIST/CORBIS

M. CAVANAUGH/AP

Research mired in Homeland Security delays

Geoff Brumfiel

Researchers hoping for financial support from the newly formed US Department of Homeland Security are increasingly anxious that their grant requests will not be answered — this year, at least.

The department was established in January by the amalgamation of several existing agencies, and was allocated \$554 million for its research and development arm for this financial year, which ends on 30 September.

Charles McQueary, undersecretary for science and technology at the department, has indicated that most of that money will initially be spent on technology, rather than science (see *Nature* 423, 106; 2003) — particularly on efforts to harness available technology for the Coast Guard, border guards and emergency first-responders.

Several researchers in subdisciplines such as chemical sensing and microbiology say they were nevertheless encouraged by their



Full speed ahead: the Coast Guard has brand new equipment — but researchers remain empty-handed.

existing funding agencies — and by officials from the homeland security department itself — to apply to it for grants.

Now they are upset that no money is forthcoming. “It seems like a bureaucratic morass,” complains Art Snow, a chemist

working on sensors at the Naval Research Laboratory in Washington DC. Snow says that he applied for funding earlier this year but has since become frustrated by the lack of communication over what kind of proposal its research wing is seeking. “I don’t know how to prepare to work with them,” he says.

One physicist at a Department of Energy laboratory, who declined to be identified, says he was told that Pentagon funding for his work in detecting chemical agents was ending, and that the new department was his best bet for future support — even though it is not yet ready to offer grants. As a result, he says, his research team is now starting to break up. And a microbiologist who works on bacterial detection at a southeastern university says that the defence department asked him to turn to the homeland security department for funds. “Everybody is saying somebody that else will do it,” he says. “And I’m out of money.”

Officials at the homeland security department, which is sitting on some 3,000 unsolicited research proposals, concede that its programmes have been delayed: so far, they say, only about half of the money due to be released by 30 September has been spent. The amount is “not as much as we would have liked”, says an official at the department’s science and technology branch who declined to be named.

But the official points out that the agency’s budget for the year wasn’t finalized by Congress until March. He says that he expects the situation to improve in the autumn.

The official said last week that the Homeland Security Advanced Research Projects Agency — a branch of the department modelled on the Defense Advanced Research Projects Agency — will ask for research proposals “within a matter of days”. It also plans to establish its first university-based research centres in November. These moves, he says, will provide more definitive points of contact for researchers hoping to get their work funded by the department.

Postdocs show independent spirit

Erika Check, Washington

Establishing just the right degree of independence is critical to forging successful relationships between postdoctoral fellows and their mentors, according to a survey at the US National Institutes of Health (NIH).

NIH officials have been probing the links between postdocs and their supervisors as part of an effort to address long-standing worries about how fairly postdocs are treated as they struggle to work their way up into permanent researchers’ positions. The issue is becoming more urgent, as more and more postdocs take several fellowships before this happens.

The survey by the NIH’s postdoctoral association of 500 of the roughly 3,000 postdocs at the agency’s main campus at Bethesda, Maryland, found that 39% of

them had done at least one fellowship before arriving at the NIH.

One finding was that those most pleased with their mentors got what they said was “the right amount” of independence. Those who felt they got too much independence, like those who thought they got too little, were less satisfied.

“Hopefully postdocs will think independently — but our experience as mentors can help bring a focus to their projects,” says Ed McCabe, physician-in-chief at the Mattel Children’s Hospital at the University of California at Los Angeles and co-author of a book on academic careers.

The issues raised in the survey will be discussed at a workshop this October led by Ruth Kirschstein, an adviser to the NIH director, which will make recommendations about what the agency should do to improve the postdoctoral experience.

Kirschstein says that the NIH has been talking to the National Postdoctoral Association, a group set up this year by postdocs in the United States. Claudina Stevenson, a founder of the association and a cell biologist who has a postdoctoral fellowship at the National Cancer Institute, says she hopes that the discussions will lead to a more proactive approach from some supervisors to postdocs’ needs.

“The first thing we want is career development, so we’re well qualified to move on to the next step,” Stevenson says. “And that doesn’t take a lot of money.”

► <http://felcom.nih.gov/mentoring/survey.html>



Parkinson's transplant therapy faces setback

Erika Check, Washington

An experimental technique that uses transplanted fetal tissue to treat Parkinson's disease is not yet ready for widespread use, according to a study published online last week.

In the study, surgeons transplanted nerve tissue from aborted fetuses into the brains of patients with Parkinson's disease. The disease destroys neurons that produce the chemical dopamine, which is required for normal brain function. The transplanted tissue is intended to replace these damaged cells.

But the average condition of the 23 patients who received the treatment did not improve significantly compared with a group of 11 who did not have it. The researchers, led by Warren Olanow of the Mount Sinai School of Medicine in New York, measured whether the treatment affected the symptoms of the disease, such as muscle tremors, speech and mental abilities (C. W. Olanow *et al. Ann. Neurol.* **54**, doi:10.1002/ana.10720; 2003).

An earlier trial, led by Curt Freed of the University of Colorado Health Sciences Center in Denver, also found that patients did not benefit as a group from fetal tissue transplants (C. R. Freed *et al. N. Engl. J. Med.* **344**, 710–719; 2001). And both trials found that the procedure caused some patients to suffer jerky involuntary movements, or dyskinesias, as a side effect. More than half of the patients in the latest trial had dyskinesias, compared with 15% of those in the earlier trial.

The results may have implications for

other research using aborted fetal tissue — and, perhaps, for future therapies that might use laboratory-grown stem cells to replace defective tissue. Some researchers fear that the negative results will reduce support for the politically controversial stem-cell therapies.

Researchers say that previous trials of fetal cell transplants for Parkinson's disease have been encouraging, as they have worked in some patients. But they feel that they cannot proceed with the technique until they work out how to prevent dyskinesias. "Unless we resolve this, it will have a tremendous negative impact on this and other issues in stem-cell research," Olanow says.

After Freed's study, scientists thought that the dyskinesias were caused by over-active neurons. But Olanow found hints that the opposite is true — poorly functioning transplants could be the cause. Olanow and others say that it is crucial to define what makes the transplants function well. Work with neurons grown from stem cells could offer clues, they add, because such lab-derived cells would contain fewer impurities than fetal tissue.

Another question raised by the latest trial is whether the transplants are being inactivated by the immune system. Olanow's team found some evidence to suggest that an immunological reaction is destroying or disabling the tissue grafts. Such evidence had not been found in previous trials.

Neurologist Olle Lindvall of the University



Clinical results cast doubt on the use of fetal cells, being collected here, to treat Parkinson's.

of Lund in Sweden, who is leading a 15-year study of fetal transplants for Parkinson's disease, remains strongly supportive of the transplant idea. But he agrees that the trials raise a number of questions that must be answered by lab work before further clinical trials are done.

"Everyone agrees that we don't yet have a method that would make fetal cell transplants competitive as a clinical treatment," he says. ■

Zoologists prime traps for California wildlife survey

Jonathan Knight, San Francisco

Zoologists in California are repeating an 80-year-old ecological survey of the Sierra Nevada as part of a scheme to measure the effect of human activity on the state's wildlife. The first phase of the project, which



Going up in the world: Californian squirrels now prefer to live on higher ground.

takes place in Yosemite National Park, is also expected to help guide the region's policy for dealing with forest fires.

From 1914 to 1920, the biologist Joseph Grinnell led one of the most extensive wildlife surveys ever conducted in the western United States. His team collected tens of thousands of animal specimens, took about 2,000 photographs, and produced some 13,000 pages of meticulous notes that are still used by biologists today.

Yosemite contains one-fifth of Grinnell's original survey sites, and this summer it once again played host to wildlife survey teams. The US National Park Service has commissioned biologists at the Museum of Vertebrate Zoology at the University of California, Berkeley, to do the work.

Whereas Grinnell relied mainly on the snaptrap — an oversized mousetrap that kills instantly — museum zoologist Craig Moritz, the head of the project, and his team are using live-trapping methods. A few animals will be kept as specimens;

the rest will be counted and released.

The team has already found some changes from Grinnell's findings. Several small mammals, such as squirrels, now seem to be living at higher elevations than they used to. One possibility is that warmer average temperatures have driven them higher; another is that fire suppression has caused a build-up of undergrowth and pushed these species uphill.

Berkeley zoologist David Wake, the team's amphibian expert, hopes that the survey will lead to the discovery of new species of salamander. Grinnell's party discovered the first Californian species, the Mount Lyell salamander (*Hydromantes platycephalus*), in Yosemite. Wake suspects that Yosemite is crawling with salamanders — but Grinnell was less interested in amphibians, and no one has looked carefully since his time.

Moritz hopes to revisit all of Grinnell's survey sites, which range throughout California from the mountains to the coast. "Yosemite will be our benchmark," he says. ■

South Korea accused of leaving birds out to dry in wetland project

Tokyo The South Korean government is in court this week to battle environmentalists over the country's largest land-reclamation project. Wildlife advocates say that the scheme ignores the plight of endangered birds.

The project, which began in 1991 and is 90% complete, involves building a 33-kilometre sea wall around the Saemangeum wetlands in the southwest of the country. Environmental groups won a court ruling in July, which called for an end to the work on the grounds that the land would not be agriculturally viable.

The government's appeal was due to begin this week, and a verdict is expected in a month's time. Critics say that the case calls into question South Korea's support for the Ramsar Convention — an agreement to protect wetlands — which it signed in 1997. The agricultural ministry counters that the region will sustain little damage, and will be able to support birds as well as grow rice.

Jeffrey McNeely, chief scientist at the World Conservation Union (IUCN), says that the project "would virtually destroy the value of the area as a fisheries production area and a wetland habitat for birds". Many birds listed as endangered by the IUCN, including the spotted greenshank and spoon-billed sandpiper, either reside in the region or use it as a migratory stopover.

Reptile study will add scale to evolutionary genetics

San Diego Snakes and lizards are set to join the ranks of species that have their genomes probed for clues to their evolutionary pedigree.

An international research consortium will examine 50 key genes in more than 145 snake and lizard species, and compare them to equivalents in the human, mouse and pufferfish genomes. Teams at Chicago's Field Museum of Natural History, the University of Adelaide, Australia, and San Diego State University in California have been awarded US\$2.4 million for the analysis.



Look ahead: reptiles such as the grass snake may give us a better view of vertebrate evolution.

Dig pinpoints scene of Donner cannibal legend

San Francisco Archaeologists believe they have unearthed the site of the dramatic final days of the Donner party, a legendary group of migrants who were stranded in the Sierra Nevada in 1846.

Historical accounts describe how 81 migrants led by the Donner family were crossing the region when early snows forced them to camp for the winter. They soon exhausted their provisions, and resorted to eating the flesh of their dead companions. About half of the group survived, including James and Margret Reed (pictured).

Julie Schablitsky, an archaeologist in Portland, Oregon, led a team that located the site based on previous excavations and artefacts unearthed by rodents. The fragments — a belt buckle, lead shot and pottery — date the site to the mid-1800s

and suggest a long-term encampment.

A bone found at the site should settle the matter, Schablitsky says. The finger-sized piece may be human, and shows hatchet marks, perhaps indicating cannibalism. Descendants of the camp's survivors have agreed to DNA testing that could match the bone to their ancestors.



By looking for genes in the scaly reptiles, biologists hope to paint a more complete picture of vertebrate evolution.

The study might even help in the treatment of snake bites, suggests project biologist Tod Reeder of San Diego State University, as antivenom from one species might be useful for treating bites from closely related ones.

Europe bans deep-sea trawling to save coral

Munich In a bid to protect an ecologically valuable deep-sea environment from damage caused by commercial fishing, the European Commission (EC) has banned the use of bottom-trawled nets on the Darwin Mounds, the largest known British deepwater coral reefs.

Discovered near Scotland in 1998 during a survey for the oil industry, the mounds make up a unique marine habitat, covering around 100 square kilometres at a depth of about 1 km. Underwater video footage showed that the reefs were already critically damaged, possibly by French trawlers, which are known to have fished in the region.

After a request by Britain's fisheries ministry, the EC, which oversees fisheries off most of Europe's coastline, has now banned trawling in the area with immediate effect. It's the first time since the revision last December of the European Common Fisheries Policy that emergency measures have been adopted to protect marine ecosystems from damage caused by fishing.

West Nile virus completes coast-to-coast journey

San Francisco The West Nile virus has reached the end of its westward march across the United States.

Health officials in California reported on 20 August that they had detected the virus in mosquitoes near the Salton Sea at the southern tip of the state. Blood tests of nearby flocks of 'sentinel' chickens have

also identified the virus.

Transmitted by mosquitoes, West Nile virus is lethal to many birds, and causes severe encephalitis in about 1% of infected humans. The latest reports provide the first evidence that the virus has established itself in California. Two previously reported human cases (see *Nature* 419, 240; 2002) are thought to have been brought in from outside the state.

The virus was expected to arrive on the west coast this summer after spreading westward since its introduction in New York in 1999. More human cases are expected to occur in California — the state's Imperial Valley near the Salton Sea is home to intensive agriculture and many outdoor workers.

Sacked Los Alamos man wins cash payout

Washington A US\$1-million settlement has been handed to a whistleblowing former employee of Los Alamos National Laboratory in New Mexico. The University of California, which manages the lab, has paid the sum to Glenn Walp, who was sacked last year after exposing theft at the lab.

Walp leaked documents to the media, revealing credit-card fraud and missing equipment, and was fired in November. His dismissal, together with that of his colleague Steven Doran, created a political storm that led to the resignations of the lab's director John Browne and his deputy Joseph Salgado (see *Nature* 421, 99–100; 2003).

Walp says he hopes that the settlement will serve as a warning to other laboratories involved in US nuclear-weapons and security programmes. He is now looking for a new job and is writing a book on his experiences at Los Alamos.

Earlier this year, Doran settled for an undisclosed sum and now works as head of security for the University of California system.



In the eye of the beholder

Does the pharmaceutical industry's future lie at the boundary between drugs and cosmetics? Or is the prospect of effective 'cosmeceuticals' a beauty myth? Helen Pearson investigates.

In 1991, dermatologist Alistair Carruthers made a surprise announcement to his colleagues. "We're going to use the most deadly agent known to humanity for the treatment of wrinkles," he told a meeting of the American Society for Dermatologic Surgery in Orlando, Florida.

That agent — a paralyzing protein extracted from the bacteria that cause botulism — was already being marketed as a drug called Botox, used to freeze the overactive muscles that cause crossed eyes. By persuading healthy people to try it off-label, Carruthers and his ophthalmologist wife Jean, at the University of British Columbia in Vancouver, Canada, had shown that it also immobilizes the muscles that furrow the brow. It wasn't until April 2002 that this treatment was formally approved by the US Food and Drug Administration (FDA). But by then, lunchtime Botox 'face-lifts' were already *de rigueur* in certain social circles.

Botox is not the only agent to have strayed across the blurred line between drugs and cosmetics. Over the past decade, a handful of drugs have found uses in flattening wrinkles, preventing baldness or shedding unwanted hair. Given the demand from a society obsessed by beauty, one might expect pharmaceutical and cosmetics companies to be falling over themselves to create a new

generation of vanity drugs — or 'cosmeceuticals', as they are known.

In truth, the action is less than intense — in large part because drug firms believe that the big bucks still lie in the treatment of disease. Although US cosmeceutical sales reached \$3.4 billion in 2002, according to the Freedonia Group, a business research company in Cleveland, Ohio, this is only a tiny fraction of the \$185.2-billion US pharmaceutical market. Even if sales of cosmeceuticals double over the next decade, as the Freedonia Group expects, that's unlikely to change the views of the big drug firms.

Face facts

The pseudo-scientific reputation of cosmetics, fear of liability should the treatment go wrong, as well as technical obstacles, are also dissuading drug companies from entering the field. But some experts predict that the firms' attitudes could change as their parched product pipelines make them increasingly desperate for commercial success. And if one company hits upon a genuine cosmeceutical blockbuster — say, a drug with no side effects that sprouts hair in all balding men — industry executives might soon be fighting to join the fray. "Their attention span could change in a flash," predicts Robert Partridge, director of communications at Dermik Lab-

oratories in Berwyn, Pennsylvania, part of Aventis Dermatologicals.

Most cosmetics firms, meanwhile, have little incentive to enter the regulatory morass of the drug business, given that they can make healthy profits from their existing range of creams and lotions. According to the FDA's definition, drugs are agents used in the diagnosis, cure, mitigation, treatment or prevention of disease, or which are intended to affect the structure or function of the body. Cosmetics, on the other hand, don't require extensive testing and FDA approval before they hit the market, because they simply alter our appearance. Cosmeceuticals can be viewed as drugs that affect our appearance, or cosmetics that change the structure or function of our bodies.

The first agent to fall into this category was all-*trans* retinoic acid, the biologically active form of vitamin A, which is now known to boost the collagen content of the skin. In the early 1980s, when Albert Kligman of the University of Pennsylvania in Philadelphia was testing all-*trans* retinoic acid as a treatment for acne in women, he found that it smoothed wrinkles caused by exposure to the sun. The drug was subsequently repackaged under the brand name Renova and approved by the FDA as an anti-ageing skin treatment. Low doses of all-*trans* retinoic acid or its precursor



Youth culture: aspects of ageing such as baldness and wrinkles are now being tackled by treatments such as Propecia (left) and Botox injections, which blur the line between drugs and cosmetics.

buster drug, the potential liabilities are also huge. Patients are normally willing to accept a small risk of experiencing side effects in treating a disease. But someone taking a drug for its cosmetic effects is likely to hold companies liable if side effects crop up. "The risk-benefit ratio is awful," says George Annas, a bioethicist at Boston University School of Public Health.

On top of this, drug companies are keen to project the image of a serious industry conquering disease, rather than one indulging in frippery. "They want to be associated with real science," says Richard Dixey, chief executive of the drug-discovery company Phytopharm, based in Godmanchester, near Cambridge, UK.

Drugs to treat obesity, on the other hand, are of interest to almost every pharmaceutical giant. Although not usually lumped together with cosmeceuticals, they are an example of the grey line between drugs that treat disease and those used for cosmetic purposes. The two leading anti-obesity prescription drugs, Hoffmann-La Roche's Xenical and Abbott Laboratories' Meridia, are only licensed to treat morbidly obese patients. But obesity researcher Stephen Bloom of Imperial College, London, says that the drugs inevitably find illicit sales to slimmers. "It's kind of self-evident," he says.

Roche and Abbott are adamant that their anti-obesity drugs are not intended for cosmetic use. Indeed, drug companies would gain nothing by aiming anti-obesity drugs at the over-the-counter slimming market — any such application would be turned down by the FDA and other national regulators because of the potential for abuse by people with eating disorders.



Despite rising demand, drug firms so far seem reluctant to pitch into the cosmetics market.

retinol now find their way into numerous over-the-counter face creams.

Most of the other well known cosmeceuticals are similarly drugs whose unexpected side effects have earned them a place in the beauty parlour. Propecia and Rogaine, for example, both of which stall hair loss, were developed to treat benign prostate tumours and high blood pressure, respectively. A recent entrant to the club is Vaniqa, the first prescription drug for removing unwanted hair. It is a topically applied version of a drug that was originally developed to treat African sleeping sickness.

Age concern

Gauging pharmaceutical companies' current interest in cosmeceuticals is difficult, because most are tight-lipped about their business strategy and research programmes. But drug giant Pfizer admits that it is actively researching new prescription compounds for cosmetic purposes. Last year, it absorbed a company called Anaderm, based in Ann Arbor, Michigan, which it had bankrolled since 1996. Anaderm is working on drugs to tackle five common cosmetic concerns: age spots, oily skin, hair loss, excess hair and sun damage. Its head of clinical exploration, Arthur Bertolino, says that two candidate drugs will go into clinical testing this year — but he refuses to reveal any further details.

Pfizer's open interest in cosmeceuticals seems to be the exception, rather than the rule. Merck, which markets Propecia, regards this product as an outlier, and says that it is continuing to concentrate on diseases such as diabetes. Industry observers say that this makes sense: not only are cosmeceuticals likely to make only a fraction of the sum that can be earned by a conventional block-

In the cosmetics industry, meanwhile, only the largest companies can afford the expensive research and clinical trials required to bring a new drug onto the prescription market. Cosmetics giant L'Oréal, for example, teamed up with the food and drinks conglomerate Nestlé in 1981 to create a dermatology-research company called Galderma, headquartered in Lausanne, Switzerland. The company is seeking new drugs to pep up ageing skin or tackle balding, says its vice-president for corporate marketing and business development, Gabriel Villada. He also declines to reveal any details.

Smooth operators

Many cosmetics firms remain wary about products that have biological activity. Nevertheless, some are exploiting the blurry definition of cosmeceuticals to bring drug-like products to market without attracting the attention of drug regulators. In the United States, if a company finds a new molecule, puts it in a pot and claims that it alters the structure and function of the skin, it must undergo years of expensive clinical trials. But if the marketing claims fudge around the cream's biological activity, it can hurdle straight into the stores. Hence the carefully worded scientific claims featured in many cosmetics adverts. "We don't want to have our products taken off the shelf," says Meagan McLellan, president of cosmetics firm Cellex-C International in Toronto, Canada.

In any case, cosmetics companies are wary of slipping potent ingredients into their creams in case they cause skin irritation. And financially, cosmetics firms seem to get by perfectly well without resorting to drug-like compounds. "There's so much money made already," says Chris Griffiths, a dermatologist at the University of Manchester, UK.

Although both pharmaceutical and cosmetics companies have business reasons to spurn cosmeceuticals, scientific obstacles also remain. The identification of hair-restoring potions, for example, has been hindered by the lack of a laboratory test bed — a scalp-in-a-dish, if you like — on which candidate molecules can be screened, says Angela Christiano, who studies the genetics of hair growth at Columbia University in New York.

But the saggy, shaggy or balding should not lose heart entirely. Entrepreneurial biotechnology companies, and pharmaceutical subsidiaries such as Anaderm, are quietly chipping away at the cosmeceutical coalface. So far, most of the drug giants are keeping their distance. "They'll only move into a new area if it has blockbuster potential," says Tom Dooley, chief executive of the dermatology-research company IntegriDerm in Birmingham, Alabama. But if IntegriDerm or another start-up firm does find the cosmeceutical equivalent of Prozac or Viagra, the big companies' attitude may change overnight. ■

Helen Pearson works in Nature's syndication news team.

Summer in Svalbard

Scientists at one of the world's most remote research outposts are tracking the air masses that swirl through the Arctic atmosphere in an attempt to understand why the ozone layer continues to thin. Quirin Schiermeier visits the Koldewey research station.



“O zonesondes, at last!” Relief shows in the face of Jens Kube as his ozone-measuring devices arrive. A small, propeller-driven Dornier aircraft has just touched down on the bumpy airstrip of Ny-Ålesund, Spitsbergen, part of the Norwegian archipelago of Svalbard. The cargo hatch opens, revealing equipment that Kube and others need to continue their research.

The delivery is one of two that arrives each week at the Koldewey station, Germany's Arctic research outpost. Each arrival brings spare parts, mail, visitors from the Norwegian mainland and excitement for the scientists who work at the isolated Arctic base, the world's northernmost permanently inhabited village, just 1,000 kilometres from the North Pole.

Temperatures are well above zero on this sunny July day. Thick inland ice covers 60% of the region's land mass, but there is open water in Spitsbergen's majestic King's Fjord, and nearby on the Barents Sea. Flowers shoot up in the green meadows surrounding the scattered wooden stations and dormitories. Outside the yellow building, where the Norwegian arctic explorer Roald Amundsen once waited in 1926 for a dirigible that would take him on a trip to the North Pole, small groups of researchers are having drinks in

the bright midnight sun.

In six months' time, things will be very different at Koldewey. During the long winter the region is an inhospitable, freezing Arctic desert, and despite the aerial grandeur of the *aurora borealis*, or Northern Lights, the station is swathed in permanent darkness. It is then that Kube's ozone-measuring equipment will begin to track a worrying problem — the continued seasonal thinning of the Arctic ozone layer. Although ozone may no longer be a political hot potato, the future status of the stratospheric ozone layer above the North Pole, and the effects of its depletion on the polar environment, is uncertain. Atmospheric scientists believe that ozone levels will remain resilient in coming decades, but they warn that we should not ignore an alarming possibility: that holes could form in the Arctic ozone layer, just as they do over the South Pole.

Clear conditions

The Koldewey station, which is headed by Kube and run by the Alfred Wegener Institute (AWI) for Polar and Marine Research's facility in Potsdam, Germany, provides a perfect arena for research. The untouched environment is free of air pollution, and the station is relatively accessible at any time of year. Germany, for example, is

Sunny outlook: Jens Kube enjoys the summer while working at his remote Arctic research base.

never more than 36 hours away; Antarctic research stations, by contrast, can be almost impossible to enter or leave during winter months. As the tail end of the Gulf Stream adds to the station's summer warmth, Italian, French, British, Japanese and Korean researchers, spanning disciplines from ornithology to glaciology, visit other nearby research stations.

There are drawbacks — Spitsbergen is polar bear country, so researchers have to carry a gun every time they leave the protection of the centre. But there are also good reasons for keeping the station open throughout the dark winter. Every day, Koldewey researchers use ground-based equipment to measure levels of ozone and aerosols in the atmosphere above the station. Balloon-borne atmospheric measurements, carried out since 1989, add extra information. Any gaps in the data would diminish their usefulness — one reason why Kube gets nervous when he runs short of measurement equipment.

Ozone measurements make up one of Koldewey's critical data sets. After all, fears over the thinning of the ozone layer have not gone away, among scientists at least. Chlorofluorocarbons (CFCs), the compounds

responsible for ozone depletion, were phased out under the 1987 Montreal Protocol, but those already in the atmosphere are likely to continue to attack the ozone layer until the middle of this century.

When CFCs reach the stratosphere, the layer of the atmosphere that stretches from about 13–50 kilometres above ground, they are exposed to solar ultraviolet (UV) radiation. This causes them to decompose, producing highly reactive chlorine atoms that break up the three oxygen atoms that form an ozone molecule. Such destruction should lessen as CFC levels fall, but this success is being tempered by another environmental threat to the ozone — the build-up of atmospheric carbon dioxide that is causing climate change.

Most of the chlorine atoms created when CFCs decompose quickly form relatively harmless compounds such as hydrochloric acid. But under certain circumstances, these compounds can re-release chlorine atoms. This reverse reaction occurs on the surface of polar stratospheric clouds (PSCs), wisps that form only in the low temperatures, around -80°C , reached during the winter in the polar stratosphere.

Layer effect

Although carbon dioxide heats the lower atmosphere, it has the opposite effect on the stratosphere. Carbon dioxide in the stratosphere reflects the UV radiation that warms the air when it is absorbed by the ozone layer. Carbon dioxide in the layer between the stratosphere and the ground also locks heat into the atmosphere at lower altitudes, preventing it from reaching higher ones. Stratospheric temperatures have fallen by around 0.5°C per decade since the 1950s, causing more PSCs to form and enhancing



Touch base: besides mail and visitors, Koldewey's twice-weekly air deliveries bring crucial equipment.

the rate at which ozone is destroyed. Moreover, if cold conditions last longer in spring, when sunlight arrives and accelerates ozone-depleting reactions, future ozone losses in the northern hemisphere may become even worse.

In the Antarctic, where local winter weather conditions isolate the stratosphere from lower, warmer parts of the atmosphere, PSCs have always been more common, and are one reason why ozone losses there have been more dramatic. Since holes in the ozone layer above the South Pole were first detected in the 1980s (ref. 1), the Arctic ozone layer has thinned every winter, but holes have not appeared.

Exactly what will happen to the Arctic ozone layer remains unclear. "We still don't know enough about the correlation between climate change and ozone depletion. It is an open question as to whether future Arctic ozone losses will get worse," says Paul

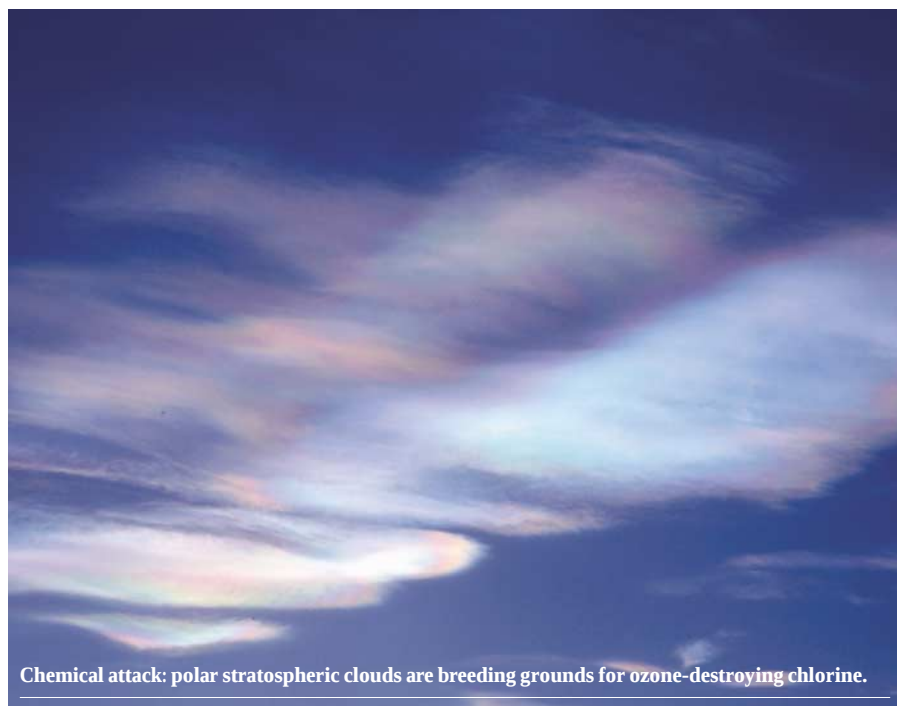
Crutzen, an atmospheric chemist at the Max Planck Institute of Chemistry in Mainz, Germany, who shared the 1995 Nobel Prize in Chemistry for his work on the formation and decomposition of ozone. "Different models arrive at very different predictions. The answer must therefore be found by *in situ* measurements."

As part of the international Network for Detection of Stratospheric Change, a global network of observation points that monitor physical and chemical conditions in the stratosphere, the Koldewey station is trying to provide those measurements. Researchers there are collaborating with colleagues at around 40 other research stations to run experiments using the Match technique, which attempts to track ozone losses in specific packets of air, about 500 km across, as they circulate around the pole. Meteorological readings and weather models are used to predict the movement of tens of thousands air parcels. Then, over a period of days, balloons are launched from participating stations under the paths of packets that were probed by previous measurements. First used in 1992, the technique allows researchers to focus on CFC-triggered ozone losses in individual packets, which are relatively isolated from the surrounding atmosphere.

Ballooning workload

As well as balloon-borne measurements, the atmosphere is probed each day — weather permitting — with light detection and ranging (LIDAR) devices, which send out pulses of laser light. The reflections of the LIDAR's vertical signals provide detailed information about PSCs and other components of the atmosphere's different layers. All raw data are checked for plausibility — a bird may have crossed the laser's path — and then transmitted to Potsdam for further analysis.

Such techniques have provided evidence that, during particularly cold Arctic winters, CFCs destroyed up to 30% of ozone in Arctic



Chemical attack: polar stratospheric clouds are breeding grounds for ozone-destroying chlorine.

regions². Although no hole has so far appeared, such an occurrence would be a more serious problem in the Northern Hemisphere than it is in the almost uninhabited Antarctic. Weather patterns in the north shift air masses over large distances and could push a hole over populated areas of Scandinavia, North America or Russia. And for every 1% decrease in stratospheric ozone, the amount of UV-B radiation — with wavelengths between 290 and 320 nanometres — reaching the ground increases by 2%, raising the risk of skin cancer by around 4% for a typical inhabitant of the affected areas³.

Researchers at Ny-Ålesund are studying the potential effects of a collapse or thinning of the ozone layer. Humans can protect themselves using sunscreen; marine Arctic ecosystems also have a safety barrier — an ice sheet — but this melts when spring arrives. At Koldewey's biology lab, scientists are investigating how marine algae react when the ice melts and increased UV radiation hits the water. Divers collect algae from King's Fjord at depths of 3–10 metres, and subject them to varying levels of artificial UV radiation in the lab.

Previous work has shown that mature algae cells can adjust to increased UV levels by expressing compounds that absorb the radiation, thus protecting their DNA and photosynthesizing apparatus from damage⁴. "Presently measured UV levels are not likely to seriously affect adult algae," says Christian Wiencke, a marine botanist and group leader of the AWI's solar UV radiation project group, who has students working at Koldewey. "But it is unclear how enhanced UV radiation might affect algae in their early developmental stages."

To address such issues, Wiencke studies single-cell zoospores, the earliest stage in the life cycle of algae. In unpublished experiments, conducted at the AWI's headquarters in Bremerhaven, Germany, he found that the spores died within 16 hours of being exposed to UV levels similar to those that penetrate the uppermost metre of Arctic seas. But he has also found that zoospores can produce



Snow place like home: the research station at Koldewey is the world's northernmost year-round town.

tannic acids — these substances had, until now, only been found in algae at the multicelled stage, where they are thought to have the potential to produce other compounds that protect cells from UV radiation.

In the long term, Wiencke and others want to develop a detailed understanding of how all types of algae will react to increases in UV radiation, and what the knock-on effects on the marine ecosystem will be. "We are beginning to understand how photosynthesis, growth and reproduction of algae respond to sudden UV stress in spring," he says. "The next step is to turn to the community level to see how algal ecosystems might be affected."

Defensive strategy

Back at the AWI in Potsdam, researchers are trying to work out whether defensive measures will have to be taken to protect Arctic ecosystems. Atmospheric physicist Markus Rex has extrapolated future Arctic ozone losses from the temperature records, trends in PSC occurrence and other data. According to his unpublished analysis, ozone losses may increase until 2010–20, and decrease only slightly by 2030. "A total collapse of the arctic ozone layer is unlikely," he says, but it is a worst-case scenario that we should not ignore."

There are great uncertainties involved in the analysis, however. Some significant processes that affect the arctic ozone layer, such as the supply of ozone from lower latitudes, and the formation of PSCs, are not yet well understood, says Rex. Because existing models fail to account for the observed large ozone losses that occurred during several cold Arctic Januaries over the past

decade, he suggests that there must be a currently unknown mechanism by which ozone can be destroyed⁵.

One theory is that ozone reacts naturally on the surfaces of PSCs, even in the absence of CFCs. Another, perhaps more likely, answer is that the chemical reactions that link CFCs to ozone destruction are stronger or faster than experiments have shown. Rex hopes that future Match campaigns, such as one currently under way in the Antarctic, will help to resolve these issues.

Back at Koldewey, the drinks continue to flow in the summer sun. Researchers there know that their geographical location gives them an advantage over their more isolated colleagues working in the Antarctic. When they finish their day's work, they can look forward to a richer social life and, in summer at least, many outdoor activities. The possibility of taking a kayak tour and encountering a polar bear, rather than a penguin, is a thrill they just have to accept.

Quirin Schiermeier is Nature's European correspondent.

1. Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
2. Rex, M. *et al.* *Nature* **389**, 835–838 (1997).
3. Van der Leun, J. C., Tang, X. & Tevini, M. (eds) *Environmental Effects of Ozone Depletion: 1998 Assessment* (UN Environ. Prog., Nairobi, 1998).
4. Bischof, K., Hanelt, D. & Wiencke, C. in *UV Radiation and Arctic Ecosystems. Ecological Studies Series Vol. 153* (ed. Hessen, D.) 227–244 (Springer, Berlin, 2001).
5. Rex, M. *et al.* *Geophys. Res. Lett.* doi:10.1029/2002GL016008 (2003).

Koldewey station

♦ www.awi-potsdam.de/www-pot/koldewey/kolnav.html

Network for Detection of Stratospheric Change

♦ www.ndsc.ncep.noaa.gov

Global Atmosphere Watch

♦ www.wmo.ch/web/arep/gaw/gaw_home.html



Bear with it: researchers who venture out into Spitsbergen's wastes must be wary of the locals.

Europe's fusion-reactor plans depend on JET

Closing the Joint European Torus may save some money, but at a great cost to physics.

Sir — In your News story “Fusion cash shortfall leaves JET grounded” (*Nature* **424**, 4; 2003), you cite Alexander Bradshaw's view that the “sensible” option for the European Union (EU) fusion programme is to close the Joint European Torus (JET), implying that it is a luxury, low-priority item.

This could not be further from the truth. JET is the most relevant device to ITER, the planned international magnetic fusion reactor, for addressing questions of how scenarios, performance limits, heat loads and stresses extrapolate with device size. As a scientist working on JET, I must point out that it is also developing key techniques that would otherwise have to await much more expensive research programmes in ITER. Examples include the development of tritium and remote-handling technologies, ion cyclotron and lower hybrid resonant

heating techniques, D–T fusion capability and exploration of effects of ‘fast’ particles (which can only be confined in a large device such as JET) in heating and changing the stability of the plasma. Thus JET is exploring the ‘new’ physics of ITER, and developing techniques and understanding to speed ITER's research programme.

In recent years JET has been redeveloped as a model of European collaboration, enabling even small scientific associations to have leading roles in key fields for ITER. Without a centrally funded EU facility, the EU fusion programme would have no coherence. We should be following the example of particle physics, pooling resources at every level in the operation of major shared facilities that most directly address the development of the field.

Closing JET may seem the easy option,

giving each country the security of having its own individual programme. A far more sensible, though challenging, option is for directors to make the case for increased fusion funding, a restructured and more centralized European programme, and to centre national resources and programmes on a few pan-European facilities.

JET remains at the forefront of the world fusion programme. With the many upgrades and new systems coming online over the next 18 months, and the vigour injected by its collaborative framework, it has one of the most diverse, dynamic and intensive ITER-relevant programmes in the world. It should be the last part of the EU programme to be closed down.

Richard Buttery

The Old Nurses Cottage, High Street, Harwell, Oxfordshire OX11 0EU, UK

Consumers don't want GM, so why use it?

Sir — The organic movement will be grateful for *Nature's* interest in our well-being (“Diversity in food technology”, *Nature* **424**, 473; 2003), but when you urge us to abandon “self-damaging dogmas”, I hope you'll forgive us for looking at your advice a little sceptically.

You are advising one of the few sectors of UK agriculture that has a real and growing market, strongly supported by consumers, to introduce a radical change in our product. We see no evidence, however, that using genetically modified (GM) crops would further the interests of organic farmers, organic food manufacturers, organic retailers or the millions of people who eat organic food in the United Kingdom.

In your Editorial you say that the Soil Association “will resist seemingly to their dying breath” the idea that GM could be as ethical as conventional plant breeding. Ultimately the decision is up to consumers. Given that people who buy non-organic food have said they don't want GM in it, it's hardly surprising that organic consumers are even more determined that GM should be kept out of organic food.

The significant areas of uncertainty described in the UK government's scientific assessment of GM crops suggest that these consumers know what they're talking about. You say that our determination to keep GM out of organic is “arbitrary and self-defeating”. Was it “arbitrary and self-defeating” when the organic sector banned the feeding of ground-up animal remains to ruminants ten years before the discovery

of BSE? This was done in the absence of any scientific evidence and solely on the basis of what you call dogma.

Thankfully the UK government has learned some lessons from past food disasters. In particular, it seems willing to listen to the market and consumers in a way that the overwhelmingly pro-GM scientific establishment in the United Kingdom finds completely impossible.

The UK government has promised to protect organic farming from GM contamination, in line with consumers' wishes (and incidentally with the European Union regulation defining organic production). As you say, there is increasing recognition of what organic farmers and environmentalists have been saying for nearly a decade: namely that coexistence of GM and organic farming may not be possible in the United Kingdom. We shall have to make a choice.

Peter Melchett

The Soil Association, Bristol House, 40–56 Victoria Street, Bristol BS1 6BY, UK

Open-source answer to bibliography problem

Sir — David M. Leslie and Meredith J. Hamilton say in their Correspondence “Multitude of reference styles delays publication” (*Nature* **424**, 127; 2003) that a standard format is needed for citation and bibliography styles. The well-established LaTeX family of open-source packages is such a system. Many journals in the physical and mathematical sciences provide

their bibliographic style files directly on their websites, reducing the problem of format management. Life-sciences journals could easily follow their example.

Leslie and Hamilton repeat a familiar objection to LaTeX: the learning curve takes away time from research work. This problem has largely been solved in the form of an open-source graphical interface to LaTeX called LyX (www.lyx.org), providing standard functionality such as cut/paste and spell-checking. New users are relieved of the time investment necessary for using LaTeX alone, yet they still derive its well-known performance benefits. The LyX interface handles standard file formats, most significantly Adobe PDF, which many journals require for electronic submission. It is therefore compatible with other tools used by authors to view, share and submit their written work.

Leslie and Hamilton discuss one standardization tool, the digital object identifier (DOI). But although there is substantial incentive for publishers to adopt DOI for increased visibility and accessibility, the benefits of simply changing long-established reference and citation styles are unclear. LaTeX/LyX may represent a more realistic solution. We hope investigators will also consider such open-source applications in the broader context of conducting their scientific work as suggested in your Editorial “In praise of open software” (*Nature* **403**, 229; 2000; doi:10.1038/35002141).

Michael C. Wendl, David J. Dooling

Genome Sequencing Center, Washington University, 4444 Forest Park Boulevard, Box 8501, St Louis, Missouri 63108, USA

The third culture

Is quantum physics, like science and literature, in a world of its own?

Quantum: A Guide for the Perplexed

by Jim Al-Khalili

Weidenfeld & Nicolson: 2003. 192 pp. £18.99

Sterling Publications: 2003. \$24.95

Frank Wilczek

C. P. Snow famously wrote of “two cultures”, scientific and literary, and of the difficulties of communication between them. To participate in the scientific culture one must be comfortable with the guiding principles of quantitative thought and strict logic, and a body of basic facts; to participate in the literary culture one must be familiar with a canon of works and expressive techniques. There are real difficulties in reconciling the two, of course, but with some effort and goodwill they can be whittled down. Scientists Freeman Dyson and Oliver Sacks, historian William McNeill and author Richard Powers are a few outstanding examples of people whose contemporary work has straddled these cultures successfully.

Problems of a different order arise when we come to “without doubt the single most important scientific advance of the twentieth century”, as the author of *Quantum* (correctly, in my opinion) describes quantum theory. Coming to terms with quantum theory requires not just quantitative thought, but the systematic replacement of unreliable intuition, derived from everyday experience, by mathematical abstraction; not just strict logic, but expanded concepts of meaning and uncertainty; not just familiarity with some basic facts, but imaginative reconciliation of apparent contradictions. Because of these difficulties, the quantum revolution has not yet had the deep impact on literary and, ultimately, common culture that its power and novelty merit.

The relatively small band of illuminati who have been initiated into the quantum culture view the world in a new and richer way. They live in a world not of four or eleven dimensions, but of infinitely many. They learn to dance “on the border between genius and madness”, enjoying “the highest form of musicality in the sphere of thought”, as Einstein variously described it. They can come to understand why chemistry works, the physical basis of quasi-magical technologies (such as microelectronics, lasers, superconductivity and magnetic resonance imaging), and the dreams of nanotechnology and quantum computing. And only they can begin to appreciate the weird but compelling concepts and algorithms that Nature uses as fundamental building blocks.

Not everyone has the time or patience

to become a licensed quantum mechanic, of course. But many more thinking people may like to be initiated into this third culture, to experience a real vision of its perspective. Certainly any aspiring philosopher (lover of insight) or theologian (interpreter of ultimate reality), or anyone who wants to imagine the technology and economy of the not-so-distant future, must become so initiated.

It is to this audience, and this purpose, that Jim Al-Khalili's new book is addressed. It is a laudable attempt to make basic concepts of quantum theory, and some of their most important applications, more widely accessible. I don't think it is the ultimate book on the subject, or even the best available, but it is a worthy effort and could serve its intended audience well.

Written in a colloquial style, the book has the feeling of a conversation with the author. You sense his sincerity and enthusiasm throughout. It makes, especially in the early chapters, for easy, charming reading. There are no equations, and many pretty pictures; if anything, too many and too pretty — quite a few of the illustrations carry little if any information content, and these may become distracting or even irritating.

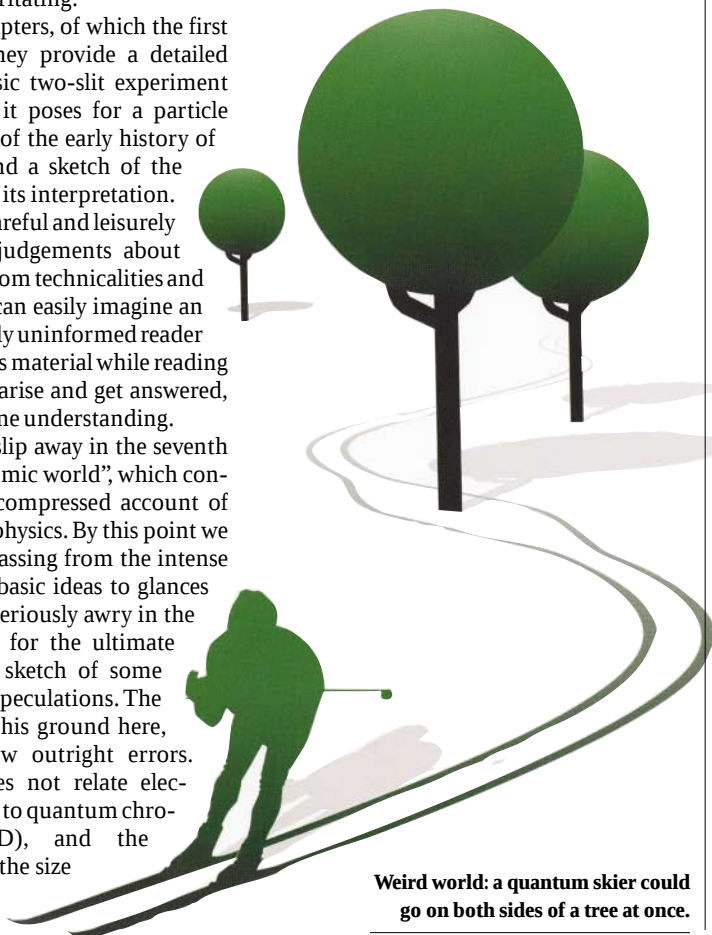
There are ten chapters, of which the first six are the best. They provide a detailed account of the classic two-slit experiment and the difficulties it poses for a particle interpretation; a bit of the early history of quantum theory; and a sketch of the ongoing debate over its interpretation. All this is done in a careful and leisurely way, with sensible judgements about when to draw back from technicalities and what to leave out. I can easily imagine an intelligent but initially uninformed reader thinking through this material while reading it, having questions arise and get answered, and coming to genuine understanding.

Things begin to slip away in the seventh chapter, “The subatomic world”, which contains an extremely compressed account of atomic and nuclear physics. By this point we have shifted gears, passing from the intense discussion of a few basic ideas to glances at many. Things go seriously awry in the eighth, “The search for the ultimate theory”, which is a sketch of some advanced ideas and speculations. The author is unsure of his ground here, and falls into a few outright errors. Supersymmetry does not relate electroweak interactions to quantum chromodynamics (QCD), and the Planck length is not the size of a quantum of space.

A more subtle and forgivable mistake, which I mention here because it is both widespread and instructive, is the claim that quantum electrodynamics (QED) provides a sufficient foundation for the physics of chemistry and biology. In reality we need to input the masses and properties of atomic nuclei if we are to account for molecular vibrations and rotations, and motion in general. These nuclear properties are governed by QCD, the theory of quarks and gluons, not by electrodynamics.

Also hyper-compressed is the ninth chapter, “Putting the quantum to work”. It provides an account of the applications of quantum physics, including semiconductor electronics, magnetic imaging, superconductivity, Bose–Einstein condensates and more, all in 30 pages of liberally illustrated large print. To be sure, a rapid survey style works better for this more grounded, less intricately layered material.

The tenth chapter, “Into the new millennium”, returns to the spirit and material of the earliest chapters. It is a tantalizing introduction to some speculative possibilities for exploiting basic quantum phenomena (such



Weird world: a quantum skier could go on both sides of a tree at once.

as entanglement and complementarity) in new forms of information processing and communication. I wish the author had gone deeper in pursuit of these logical outgrowths of his core themes, instead of scampering to touch all bases.

For seekers of the third culture, I'd still recommend Richard Feynman as the best guide. Try *The Character of Physical Law* (MIT Press, 1967), the early chapters in each volume of *The Feynman Lectures on Physics* and, for the ambitious, *QED: The Strange Theory of Light and Matter* (Princeton University Press, 1986). ■

Frank Wilczek is at the Center for Theoretical Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue 6-305, Cambridge, Massachusetts 02139-4307, USA.

Taming the world?

Our Own Devices: The Past and Future of Body Technology

by Edward Tenner

Knopf, 2003. 336 pp. \$26.

Howard P. Segal

The unintended effects of technological developments may be more harmful than beneficial. This was the message — familiar to scholars but not to the general public — that Edward Tenner discussed in original and amusing ways in his previous book *Why Things Bite Back: Technology and the Revenge Effect* (see my review in *Nature* 382, 504–505; 1996). In his latest book, *Our Own Devices*, Tenner examines the ceaseless interplay between technological devices and technique, or our ability to use them.

Most students of technology, he observes, concentrate on the devices themselves; their applications are usually harder to analyse. Yet Tenner also rejects the influential arguments of social critic Jacques Ellul that technique, broadly defined, has steadily shaped modern “technological society” and that its influence is inescapable. (The original French title of Ellul's most influential book, *The Technological Society*, was *La Technique*.)

Tenner focuses on several “body technologies” used in everyday life, and moves literally from toe to head: sandals, training shoes, chairs for office and home, musical keyboards, text keyboards, babies' bottles, glasses and helmets. Overall, these technologies have provided us with a more comfortable, efficient and protected existence as we have

modified the natural world — which is how the earliest technologies began. Invariably, though, there are those “revenge effects”. But *Our Own Devices* goes further than the first book in demonstrating how far body technologies have defined us as human beings.

Babies' bottles are usually the first technology that most of us encounter, but the technique of nursing must still be mastered by most children and mothers who seek to utilize both bottles and breasts. Regular use of rubber bottle nipples often diminishes infants' use of jaws and tongues to breastfeed. And too much baby formula may increase the prospects for certain adult ailments.

To take another example, most of us remain indebted to Roman sandals for the basic model of walking comfort in warmer climates. But comfort is not always a necessity, as the world's billion barefoot inhabitants could attest. For the rest of us, though, sandals — and shoes — make our feet too sensitive to manage without them. Yet cultural differences may dictate our particular choices, so Japanese men wearing geta (raised clogs) walk differently from their Western counterparts.

The role of literacy is critical to Tenner's analysis of glasses, chairs and keyboards, all of which affect us both physically and intellectually. Not surprisingly, glasses go hand-in-hand with mass literacy — but so does myopia. Ironically, glasses may cause children's eyes to grow differently, and so increase myopia as much as correct it. Meanwhile, a few straight-backed chairs for monks and clerks existed before ordinary people learned to read and write, but only with mass literacy did such chairs — and later reclining chairs — become fundamental to those enterprises. Reclining chairs, however, often weaken our back muscles, causing us to recline still more to be comfortable.

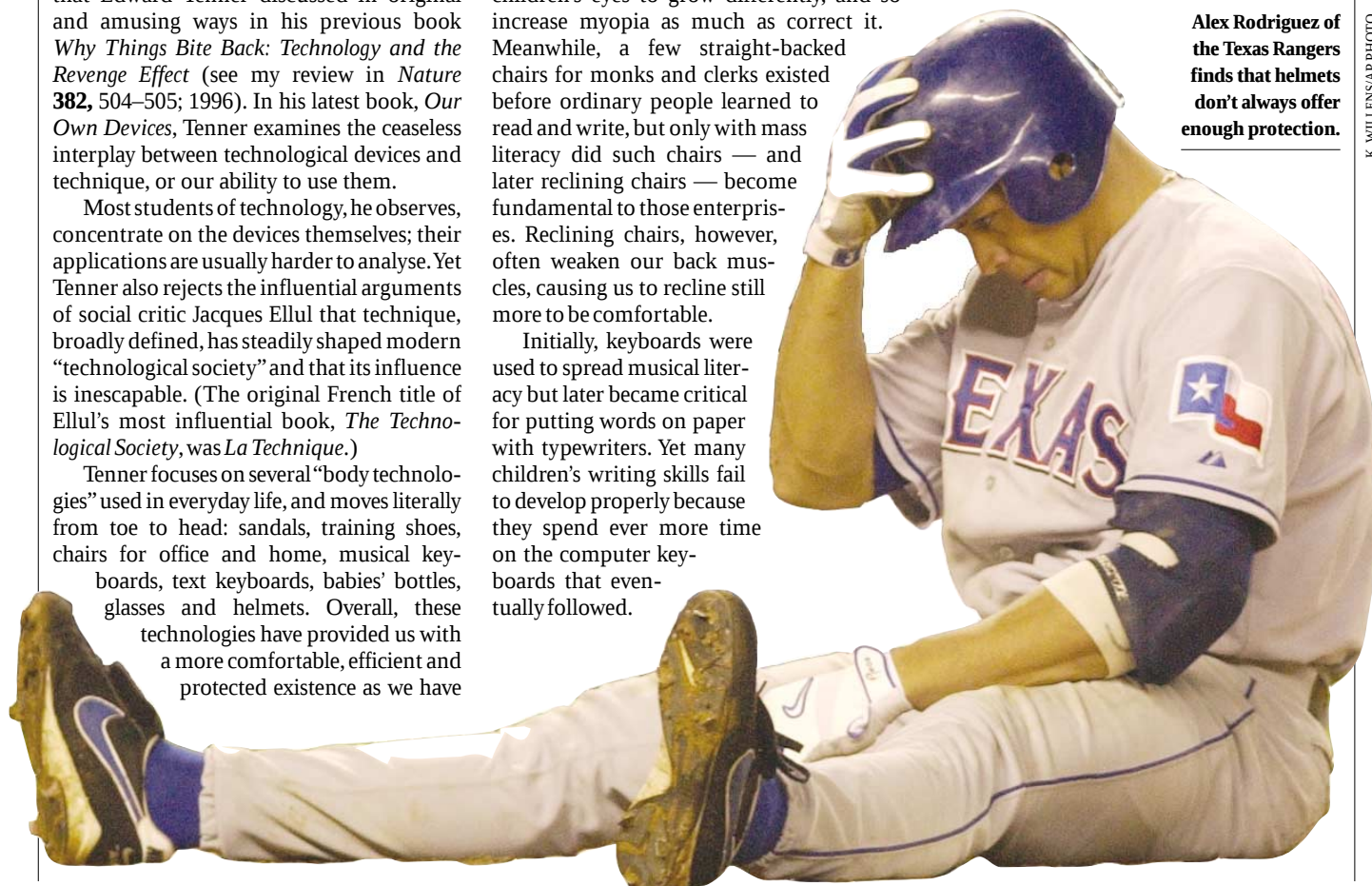
Initially, keyboards were used to spread musical literacy but later became critical for putting words on paper with typewriters. Yet many children's writing skills fail to develop properly because they spend ever more time on the computer keyboards that eventually followed.

Helmets provide one of Tenner's most intriguing examples. They were invented in classical times, but their success prompted the development of more powerful weapons and, Tenner claims, started the arms race. Adapted for sports, stronger, and presumably safer, helmets have ironically led to ever more violent contact and greater injuries.

However old in origin, all of Tenner's examples in their current forms reflect the sea change in conceptions about technology, in the West at least, over recent decades. We seem ever more concerned with technology providing small-scale fulfilment of individual needs and desires than with providing large-scale solutions to widely acknowledged social problems — from insufficient energy, communications and transportation systems to excessive poverty and population. Body technologies, rather than space programmes or nuclear power plants, have become our primary association with technology and technique. Tenner is not concerned with this trend, and seems happy to instruct and entertain in a non-ideological, often wry manner.

Tenner then briefly discusses the proliferation of mobile phones and palm pilots. Despite the dominance of the index finger, which represents ‘authority’, Tenner predicts the rise of the thumb: “when extended in the almost lost art of hitchhiking, the thumb shows the right attitude toward the future,

Alex Rodriguez of the Texas Rangers finds that helmets don't always offer enough protection.



K. WILLENS/AP PHOTO

open and cooperative but with a firm sense of direction."

Our Own Devices is even more insightful and provocative than Tenner's earlier book in illuminating how contemporary technology changes us as much as we change it. Tenner has become a worthy successor to such luminaries as business philosopher Peter Drucker, social critic Lewis Mumford and historian Lynn White in connecting technology's past, present and future. ■

Howard P. Segal is in the Department of History, University of Maine, Orono, Maine 04469-5774, USA.

The spark of Enlightenment

Volta: Science and Culture in the Age of Enlightenment

by Giuliano Pancaldi

Princeton University Press: 2003. 384 pp. \$35, £24.95

Lucio Fregonese

Alessandro Volta (1745–1827) was a leading natural philosopher who made significant contributions to fields as diverse as electricity, chemistry and pneumatics. His famous electrical battery contributed decisively to the rise of modernity.

Pancaldi's book is a scientific biography of Volta within the context of Enlightenment culture. The author examines central aspects of Volta's biography, including his education and religious attitudes, and traces a lively picture of the social and scientific networks, both local and international, in which he operated. The sources used include Volta's correspondence, the main Italian scientific journals and the surviving information on the many journeys he made to meet and talk with other scholars.

The focus is mainly on Volta's work on static electricity (1765–1787) and the advanced stages (1796–1799) of the long path that took him to the invention of the battery. Ideally, these two achievements should be considered alongside his work in other scientific fields, but the partial state of our knowledge justifies this selective approach.

Pancaldi's interpretation of Volta's static electricity rests on a sharp polarization, which gives prominence to "experiment and apparatus over theory, instrumentalism over deep philosophical commitment, efficacy and public recognition over private achievement". Not all historians or philosophers of science will agree with such radical epistemological and historiographic assumptions.

In discussing the portrait of Volta in old age shown here, Pancaldi singles out three of the elements — theory, experiment and apparatus, and public recognition — that

appear in his pairs of polar opposites. The book in Volta's hand represents theory; the battery and the *condensatore* (the first version of the modern capacitance electroscope) in the foreground represent his electrical instruments; and the ribbon of the *Legion d'Honneur* on his jacket collar represents public recognition.

Pancaldi interprets Volta's invention in 1775 of the electrophorus (a new electrostatic generator that uses what we now call electrostatic induction instead of friction) as having been inspired not by theory but by the desire for an improved and effective apparatus, aimed at gaining social and scientific recognition. He claims that Volta simply imitated and improved on a similar device described years earlier by the German physicist Franz Aepinus. In about 1778, Volta is presented as having finally managed to define a sound theoretical framework, which was however "midrange", because it built on concepts (capacity, tension and actuation) that, according to Pancaldi, "stemmed directly from laboratory practice". Aepinus and two other important physicists working with electricity, Henry Cavendish and Giambattista Beccaria, are indicated as likely sources of inspiration. Volta's invention (1780) and conceptualization (1782) of the *condensatore* are rooted in his newly defined handling of electricity. Volta's later extensive efforts to quantify various electrical quantities, especially tension, are connected to the general "quantifying spirit" of the Enlightenment.

Pancaldi provides important insights into Volta's path to the battery. He shows that Volta's detection of contact electricity emerged in the context of the more general "hunt for weak electricity" pursued at the time. He also argues convincingly that Volta's creation of the battery was probably catalysed by a paper in which the English chemist William Nicholson had described an apparatus suggested to work like the electrical organ of the torpedo fish. The battery is accordingly presented as a largely "unintended consequence" of a process of "competitive imitation" with Nicholson to reproduce the functioning of the animal's organ. The relative merits of theory and practice are evaluated just as for Volta's static electricity. Experiment, laboratory-based conceptualization and instrumentalism are placed at the core of Volta's enterprise. His deep theoretical commitments are correspondingly given less weight.

Pancaldi offers new results on the way in which Volta presented the battery on various occasions and on how it was received in different contexts (the English and French press) and by various scholars (Nicholson, Etienne-Gaspard Robertson, Johann Wilhelm Ritter, Hans Christian Ørsted and Jean-Baptiste Biot). The battery found many heterogeneous uses and interpretations, which Pancaldi sees as further evidence of "unintended



Fulfilling his potential: Alessandro Volta valued theory, experiments and public recognition.

consequences". Another chapter focuses on the celebration of Volta, including iconography and the centennial commemoration of his death in fascist Italy.

Pancaldi's final image of Volta is of an instrumentalist with manipulative and eclectic tendencies, characteristics that Volta certainly had. But Volta's natural philosophy seems to require further clarification. Volta himself denied that Aepinus' device was the inspiration for his electrophorus, claiming instead that he produced it with the highly theoretical goal of confirming his own conceptualization of a very complex area called "vindicating electricity". As Pancaldi points out, Volta continued to allude to his deep theoretical views as a prime mover of basically all his electrical achievements, including his controversial hypothesis (1792) of contact electricity. As evidence of this, Volta repeatedly referred to his very first electrical paper (1769), a difficult work that remains the only systematic theoretical exposition he ever produced in this field.

From the portrait reproduced here, Volta seems indeed to reassert a deep theoretical commitment, with his natural philosophy as a whole (the book) indicated as the origin of his electrical instruments. Conversely, he can be seen as claiming that the instruments speak in favour of his own natural philosophy. It seems that further enquiry and reassessment are needed for both Volta's natural philosophy and its links with the instruments he produced. ■

Lucio Fregonese is in the Museo per la Storia, University of Pavia, Strada Nuova, 65, I-27100 Pavia, Italy.

Hunting antimatter

Summarize yourself in the form of a title of a paper in Nature.

A minuscule part of the Universe trying to understand the rest.

What was your first experiment as a child?

The first in an (unpublished) series of experiments that explored the effects of gravity while riding a tricycle down the stairs at the age of three. The subsequent somersault answered my initial question of whether the stairs were too steep with a clear 'yes'.

What book has been most influential in your scientific career?

The *Feynman Lectures*, which I encountered in my first year of physics. Although I did not fully understand the beauty of Feynman's approach right from the beginning, the lectures inspired me to think about physics from an intuitive and elementary point of view.

Who has been the most important mentor in your career?

Rafael Armenteros. When I came to CERN, he was one of the world's leading authorities on anti-proton physics. He was quiet, modest and an excellent scientist. His simple but deep way of thinking taught me a lot about research.

What single scientific paper or talk changed your career path?

At the CERN Summer Students' Lectures in 1976, Haim Harari gave a talk on the (then) new 'standard model', based on six types of quarks and leptons and three interactions. The relative simplicity and the enormous predictive power made me realize that this was a great step forward. Compared with the previous dark age and its ever-increasing zoo of 'elementary' particles and complicated field theories, it was like seeing the light at the end of a long tunnel.

What overlooked or underrated discovery really changed the science in which you work?

Electron cooling of particle beams. Invented by Gersh Budker and Alexander Skrinisky in Novosibirsk in 1967, it has proved to be a crucial technology for many low-energy storage rings around the world. It is also, in a slightly modified version, very useful for cooling trapped particles.

What literary character would you employ as a postdoc?

Harry Potter. Fixing the apparatus sometimes needs magic.

What music heads the playlist in your car or lab?

Rhythm and blues or hip-hop in the car. But no music in the lab, because it turns me off work.

What's your favourite conference destination, and why?

London. Great conference facilities, but easy escape routes to restaurants and entertainment in the evening.

Whose graduate student would you most like to have been?

Enrico Fermi, because of his unique creativity both in theory and experiment.

What book is currently on your bedside table?

Longitude by Dava Sobel.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Marilyn Monroe. The conversation would certainly never become boring.

Where and when would you most like to have lived or worked?

Göttingen and Copenhagen in the 1920s, hiking and sailing with Bohr and Heisenberg, and witnessing the development of quantum theory.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Listen carefully. I might find out what others really think of me.

What's the best piece of advice you've received?

Don't fix it if it's not broken.

What do you most dislike about having research published?

The book-keeping that is needed to establish a solid chain of arguments.

The Internet is the bane of scientists' lives because...

...every time I ask a simple question, people just tell me: "But I put the answer on the web ages ago."

Is there a 'tyranny of reductionism' in how scientists are trained today?

To make an impact in frontier research, many young scientists need to focus on a very narrow field of science or applied technology. Unfortunately, they often stay in this highly specialized area when they get older. Crossing the borders between different fields of research should be rewarded, not punished.

What's the one thing about science that you wish the public understood better?

Scientific research is a long process of systematic trial-and-error, based on models that are mostly proved wrong.



Rolf Landua

Rolf Landua works on the production of anti-hydrogen atoms at CERN, and hopes to understand better why the Universe contains so little antimatter.

What do you do to relax?

I practise a Vietnamese martial art called vovinam vietvadao, a mixture of kung fu, tae kwon do, ju jitsu and karate. Learning many complex moves keeps the brain entertained while the body is working out. Highly recommended for fitness and relaxation after a stressful day.

Which field in science (apart from your own) deserves more funding, and why?

The development of medical applications from new technologies used in frontier physics experiments. This would provide the motivation to exploit the potential of new technologies for the benefit of society at large.

What single discovery, invention or innovation would most improve your life?

An intelligent search engine that gives me exactly the information I want.

What would you have become, if not a scientist?

A DJ in a fashionable nightclub or a fitness trainer in a model agency.

What music would you have played at your funeral?

The Show Must Go On by Queen.

What's just around the corner?

Understanding dark matter; efficient cancer therapies without serious side effects; solar- and fuel-cell technology in our daily life; household robots with human appearance and suitable for diverse purposes; an all-in-one pocket-sized communication/entertainment/computing device that will make portable computers and mobile phones look like stone-age tools.

Genome sequences from the sea

Jed Fuhrman

Despite their diminutive stature, phytoplankton have a huge global influence. The genomes of four strains of phytoplankton have now been completely sequenced, revealing their genetic adaptations to distinct marine niches.

With the sequencing of microbial genomes now almost routine in some circles, one could be forgiven for feeling a little jaded. It is possible to lose sight of just how much can be learned from such an exercise. But the reports by Rocap *et al.*¹ and Palenik *et al.*² in this issue, along with a paper by Dufresne *et al.*³ in *Proceedings of the National Academy of Sciences*, powerfully demonstrate how genomic studies can lead to a new understanding of biodiversity, ecology, biological efficiency and biogeochemistry.

About half of global photosynthesis and oxygen production is accomplished by single-celled planktonic organisms (phytoplankton) that live in the top layer of the ocean, where enough light penetrates to support their growth. The most plentiful of these are the cyanobacteria — tiny, chlorophyll-containing phytoplankton that have no membrane-bound nucleus. There are two basic types. The *Synechococcus* strains have a diameter of about 0.9 µm and were discovered to be abundant in seawater in 1979 (ref. 4). With a diameter of about 0.6 µm, meanwhile, the *Prochlorococcus* strains (Fig. 1) are the smallest of all the phytoplankton; they are also the most abundant, yet were discovered only 15 years ago⁵.

Now the complete genomes of three strains of *Prochlorococcus*^{1,3} and one strain of *Synechococcus*² have been sequenced and analysed. The results are remarkable for what they show not only about the differences between these close relatives, but also about

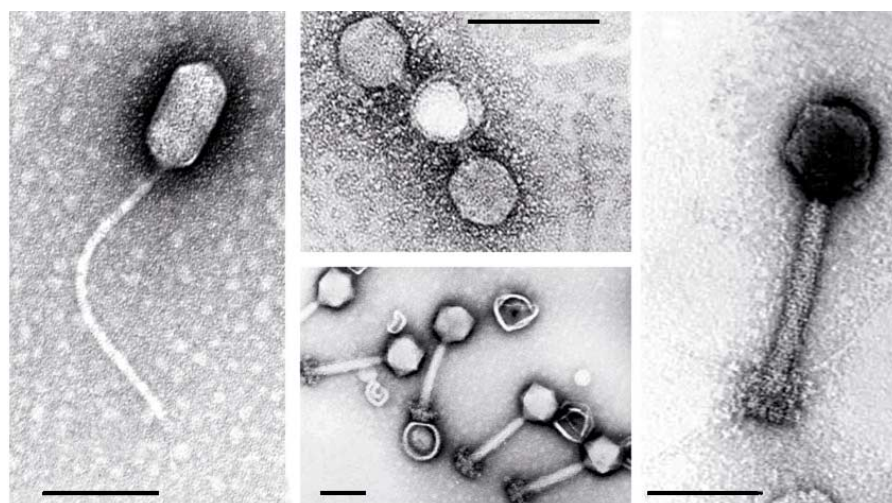


Figure 2 Marine cyanophages. Cyanophages are viruses that infect cyanobacteria; those shown here infect *Prochlorococcus*. The newly reported whole-genome sequence data^{1–3} suggest that cyanophages might be an important factor driving the evolution of cyanobacteria. Also in this issue, Sullivan *et al.*¹¹ describe the first isolation of *Prochlorococcus*-infecting cyanophages, with some phages being capable of infecting multiple strains of *Prochlorococcus*, and even of crossing to the closely related *Synechococcus*. Cross-infection is a possible mechanism of gene transfer between different hosts. Scale bars represent 100 nm. Left, siphovirus P-SS2; top, podovirus P-SSP9; right and bottom, myoviruses P-RSM2.

the extent to which some strains economize on DNA. Indeed, as Rocap *et al.*¹ and Dufresne *et al.*³ describe, two of the *Prochlorococcus* strains have genomes so small — about 1.7 million base pairs — that they might represent the minimal genome of an oxygen-generating, photosynthetic organism.

Why economize on DNA? A genome represents the complete genetic repertoire of an organism: it contains specifications for all the machinery needed to create the organism and to regulate its operation, growth and reproduction. Control of genome size involves a trade-off between efficiency and versatility. A small genome reduces the amount of extra 'baggage' that must be maintained and propagated, but also limits an organism's ability to exploit different resources. Larger genomes provide for this possibility, and also permit back-up should a gene be damaged or lost. But they require more material and energy to maintain.

For some *Prochlorococcus* strains, it seems that the minimalist approach to genomes works best^{1,3}. *Prochlorococcus* is a photosynthetic 'autotroph' (an organism that uses inorganic sources of carbon), and its genome must encode proteins that allow all cellular components to be synthesized from simple

inorganic compounds, with power for all this coming from sunlight. In the top layer of the sea, nutrients such as nitrogen and phosphorus are often extremely dilute. So the two-pronged strategy of being physically small (reducing cellular requirements) and of economizing on DNA (which contains both nitrogen and phosphorus) can clearly be very successful — although there are other strategies, of course.

The tiny genomes now studied belong to the high-light-adapted *Prochlorococcus* strain MED4 (ref. 1) and the very-low-light-adapted strain SS120 (ref. 3), which might be able to get away with such genomic paucity in part because of the stability of their environments. They grow optimally in the bright (near-surface) and dim (deeper) portions of the water column, respectively; SS120 cannot photosynthesize at the light levels that are optimal for MED4 (ref. 6). The different kinds of *Prochlorococcus* can do well when specialized for certain depths because the warm ocean water where they are typically found is usually stratified by density — so different depths rarely mix with each other. Resources such as light and nutrients are also stratified (see Fig. 1 of ref. 1, page 1043).

For the other strains now studied, larger genome size probably relates to the use of

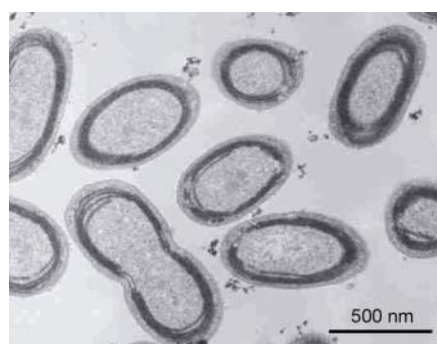


Figure 1 *Prochlorococcus*: the smallest of the marine phytoplankton, and the genus with the smallest known genomes of any oxygen-generating photosynthetic organism. This electron micrograph shows strain MIT9313, one of the strains whose complete genome sequence is now reported^{1–3}.

M. SULLIVAN/MIT

a broader variety of resources. The third *Prochlorococcus* strain, MIT9313, has adaptations that seem to place its optimum depth between that of MED4 and SS120 in the water column. Rocap *et al.* find that it also has a larger genome (2.4 million base pairs) and more genomic versatility, perhaps enabling it to exploit the diverse resources available in the 'transition' zone where it lives. According to Palenik *et al.*², the *Synechococcus* strain WH8102 also has a larger genome (again, about 2.4 million base pairs). This organism is known to fare particularly well under conditions of upwelling and vertical mixing, where individual cells can be exposed to a large variety of nutrient and light conditions. Its bigger genome and greater number of genes presumably give it the ability to acclimate. For instance, the sequences imply that *Synechococcus* strain WH8102 can use ammonium, nitrite, nitrate, urea, cyanate, amino acids and peptides as sources of nitrogen. *Prochlorococcus* MIT9313 can use all except nitrate and cyanate, MED4 only ammonium, urea, cyanate and peptides, and SS120 only ammonium and amino acids.

The remarkable partitioning of resources such as light and nutrients between these close cyanobacterial relatives beautifully illustrates the question of the meaning and significance of the term 'biodiversity' when it comes to bacteria. Although animal and plant diversity is continually in the news, with evidence mounting about serious losses, microbiologists are often hung up on simple questions such as how to measure diversity or define a species. Because most bacteria lack taxonomically useful morphological features, classification studies now rely heavily on molecular sequences. New molecular biological techniques allow microbial diversity to be determined directly from natural communities, without needing to cultivate bacteria in the laboratory (which is often problematic)⁷.

But it is not easy to interpret the results of sequence-based diversity studies in a classical ecological framework, or to reconcile the results obtained by different methods. For example, one recent study used a bioinformatic analysis of sequence databases to estimate that marine bacterial species number only a few thousand⁸. In contrast, another used abundance curves to calculate that there may be as many as two million distinct taxa⁹. Interestingly, both sorts of analysis would have lumped most of the cyanobacteria sequenced here¹⁻³ into a single taxon or perhaps a single species, because such analyses tend to use a criterion of more than 97% sequence identity in the small subunit of ribosomal RNA to define a species or taxon (and the three *Prochlorococcus* strains fit that criterion). Clearly that would be a mistake, given what we know about the physiological and, now, the genomic differences. And that

suggests that previous estimates of the total diversity of marine bacteria are rather low.

Can genomes provide clues to how biodiversity is created or maintained? Indeed they can: the new studies¹⁻³ tell a story of rapid evolutionary adaptation, including gene loss and a significant transfer of genes from other bacteria. Of particular note are changes in genes controlling the cell surface; these affect susceptibility to viruses (Fig. 2), and possibly recognition by protists that feed on the cyanobacteria — which is consistent with ecological investigations that implicate viruses or grazers as important selective agents in bacterial evolution¹⁰. Ironically, several of the gene-transfer events might themselves have been mediated by viruses, as demonstrated by the flanking of transferred genetic regions by viral-like integrase sequences²; integrases are enzymes that insert viral nucleic acids into a host genome.

Yet more surprises hide in these data. For example, as mentioned above, *Prochlorococcus* MED4 maintains genes that allow cyanate to be used as a nitrogen source. But until now, biological oceanographers have not thought of cyanate as being a nitrogen source of any significance. The same applies to phosphonates (molecules with C-P bonds) as a source of phosphorus. Then there is the mystery of how these organisms can perform some enzymatic functions even

though they apparently lack genes similar to those that perform these functions in other organisms. For instance, there is no recognizable gene for carbonic anhydrase, which converts bicarbonate to CO₂ for photosynthesis. In fact, about one-third of each of the newly sequenced genomes bears no relation to anything previously identified. These genes are a veritable treasure-trove that should point to novel physiological and ecological phenomena. Who would have thought 20 years ago that marine ecologists would be eagerly awaiting gene sequences? ■

Jed Fuhrman is in the Department of Biological Sciences and the Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089-0371, USA.
e-mail: fuhrman@usc.edu

1. Rocap, G. *et al.* *Nature* **424**, 1042–1047 (2003).
2. Palenik, B. *et al.* *Nature* **424**, 1037–1042 (2003).
3. Dufresne, A. *et al.* *Proc. Natl Acad. Sci. USA*; published online 13 August 2003 (doi: 10.1073/pnas.1733211100).
4. Waterbury, J. B., Watson, S. W., Guillard, R. L. L. & Brand, L. E. *Nature* **227**, 293–294 (1979).
5. Chisholm, S. W. *et al.* *Nature* **334**, 340–343 (1988).
6. Moore, L. R. & Chisholm, S. W. *Limnol. Oceanogr.* **44**, 628–638 (1999).
7. Pace, N. R. *Science* **276**, 734–740 (1997).
8. Hagstrom, A. *et al.* *Appl. Env. Microbiol.* **68**, 3628–3633 (2002).
9. Curtis, T. P., Sloan, W. T. & Scannell, J. W. *Proc. Natl Acad. Sci. USA* **99**, 10494–10499 (2002).
10. Fuhrman, J. A. *Nature* **399**, 541–548 (1999).
11. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. *Nature* **424**, 1047–1051 (2003).

Turbulence

Suddenly it's chaos

George Pickett

Injecting vortices into a rotating sample of superfluid helium-3 shows a sudden switch from smooth to chaotic behaviour, and throws light on turbulence — one of the last unsolved problems of classical physics.

Leonardo sketched it. Hydrodynamicists model it. Turbulence in fluid flows is universal. Not only does it impinge directly on human existence, but it affects physical behaviour on all scales, from the dynamics of galaxies to particle motion inside the atomic nucleus. Turbulence is clearly important and we know a great deal about it, but there is no satisfactory theory for it. One way forward is to attack the problem in its simplest form, as A. P. Finn *et al.* have done in their experiments reported in this issue (*Nature* **424**, 1022–1025; 2003).

This group is studying vortices in the simplest of fluids, superfluid helium. In a normal liquid the constituent particles are distributed over many possible quantum-mechanical energy states. Constant collisions ensure that the behaviour on a microscopic level is a frantic redistribution of particles among the available states. To model that, we would need one set of

equations to track the career of every constituent particle, and that is quite beyond us. In contrast, when liquid helium is cooled to temperatures close to absolute zero, a large fraction of the constituent particles occupy the lowest available energy state and so they all behave alike — they march in step, rather than running around at random. Therefore there is no friction. The liquid is a superfluid.

That makes for simplification, but it also presents a problem. To investigate turbulence, the natural place to start is by imagining a rotating bucket of superfluid. In fact, a superfluid cannot be set into simple rotation. To do so, we would have to force it into 'solid body rotation', for which the tangential velocity increases linearly from the centre. But there is no way in which we can devise a wavefunction for the superfluid, describing the state of every constituent particle, that will achieve this. It is easy to create a wavefunction in which the velocity falls as we go

away from the centre, as it does in a tornado, but not the other way round. Superfluids are hence said to be inherently 'irrotational'.

However, miniature tornados — vortices — can exist inside a superfluid. In a vortex, all rotation is concentrated along its core axis, and the liquid outside is irrotational. To picture this, imagine placing a cork in a rotating bucket of water: the cork rotates about its own axis once per cycle; the liquid rotates everywhere. But if you put a cork in a bucket with a vortex of liquid running out of a hole at the bottom, the cork will circle the vortex but will not rotate. It only rotates if trapped on the axis. In the superfluid we gain a further simplification because, as the flow around the core is quantized, vortices in superfluids are all alike. Rotation in a superfluid can be simulated on a large scale by a pattern of vortices; so, paradoxically, large turbulent patterns, in which we would expect violent rotation of the liquid, can be set up in a superfluid by creating an enormous numbers of quantized vortices. On a large scale the liquid seems to rotate, but on a microscopic scale it does not.

The experiment done by Finne *et al.* probes how vortices in the superfluid attain equilibrium in a rotating chamber. This is not a trivial question. The superfluid does not sense the rotation of the container. However, a fraction of the particles in the system are not in the ground state; instead they behave normally and rotate with the container. The liquid can thus be considered to be a mixture of superfluid and normal fluid: at absolute zero all the liquid is superfluid; with increasing temperature the fraction of normal fluid increases, reaching 100% at the transition temperature where superfluid behaviour vanishes (about 0.001 K for helium-3). It's as though the normal fluid 'informs' the superfluid of what the container is doing: the vagaries of this 'conversation' are the subject of Finne and colleagues' work.

The authors set a container of superfluid helium-3 into rotation, then, by various methods, nucleate a small vortex loop in the system, tracking its progress to equilibrium by nuclear magnetic resonance. Figure 1 shows a simulation by Finne *et al.* of the behaviour of the superfluid at 'high' temperature — 80% of the transition temperature — and low temperature (40% of the transition temperature). At high temperature, there is plenty of normal fluid in the mix. The container rotates about its vertical axis and the equilibrium distribution would be a uniformly spaced array of vertical vortices. When a vortex loop is nucleated at the base, the interaction with the normal fluid damps any tendency of the vortex to oscillate, and simply pulls it smoothly into its vertical equilibrium position. Each subsequent nucleation adds one vertical vortex to the pattern until equilibrium is achieved. At low temper-

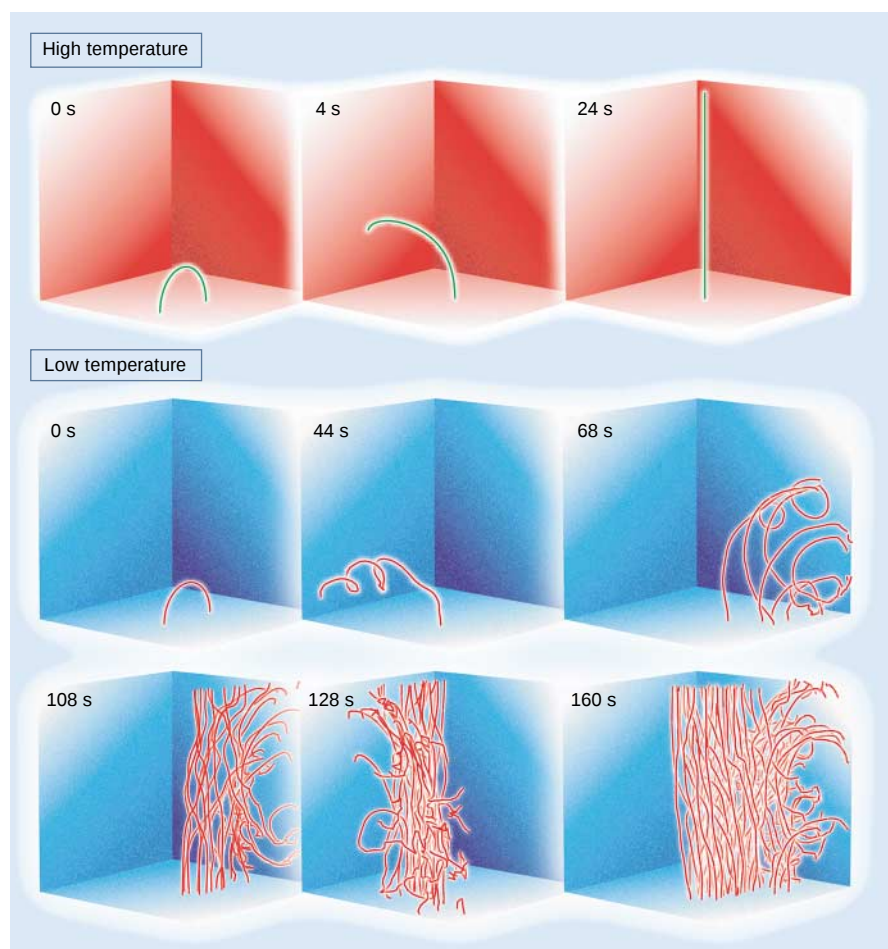


Figure 1 Superfluid turbulence. A selection of frames from simulations by Finne *et al.* (whose experiments are reported on page 1022 of this issue) shows the development of vortices inside a rotating superfluid of helium-3. At 'high' temperature — meaning at 80% of the transition temperature at which helium-3 becomes a superfluid — a vortex loop evolves smoothly into its vertical equilibrium position within 24 seconds. But low-temperature behaviour (at 40% of the transition temperature) is quite different. A single vortex loop evolves chaotically, eventually filling the container volume with a tangle of vortices. Finne *et al.* find that the switch from smooth to chaotic behaviour happens over a surprisingly narrow temperature range. (For technical reasons the simulation is made for a container of square cross-section.)

atures the behaviour is entirely different: the tenuous normal fluid only weakly transmits information about the container, and the damping forces are correspondingly smaller. Now, when a vortex loop is nucleated, it can flail about essentially unrestrained. When sections cross they can recombine, forming new loops, and the whole volume fills with a tangle of vortices. Damping forces gradually bring this tangle into the equilibrium array hinted at in the final frame of Fig. 1. So a single nucleation event can provide the full equilibrium array. This behaviour is much as we would expect. The completely unexpected observation is that the switch from high-temperature smooth behaviour to low-temperature chaotic behaviour occurs over a very narrow temperature range indeed — only 5% of the whole range of superfluid behaviour. That is a real surprise.

Lest one should think that observing the vortex lattice in rotating helium-3 at microkelvin temperatures is a rather solitary

interest, remember that this system mimics the rotating neutron superfluids inside neutron stars. The rock-steady precession of these objects is occasionally interrupted by a glitch in rotation frequency, a 'starquake', caused by the vortex lattice in the core reorganizing itself as the spin rate of the star gradually slows down. Finne *et al.* have shown that what should be simple vortex behaviour in the simplest of liquids still gives rise to unexpected phenomena, conundrums that must be solved on the road to understanding superfluid vortex dynamics. But once armed with that knowledge we can make a more solid attack on understanding turbulence in general, examples of which pervade all of modern life — from plasmas, the weather, lubrication, aerodynamics and blood flow to the water running out of your bath.

George Pickett is in the Department of Physics, Lancaster University, Lancaster LA1 4YB, UK. e-mail: g.pickett@lancaster.ac.uk

Of mice and mentality

Steve Blinkhorn

Evidence of a general learning ability in mice — that there is a good correlation between an individual's performance in tasks that make different processing demands — suggests a parallel with humans.

Why isn't everyone a genius? Argue it whichever way you like, people manifestly differ in the ease with which they learn, and, whatever the reason, we have to provide for these differences in education, training and employment. But for the study of the elementary processes of learning, psychologists have traditionally looked to laboratory animals: rats, mice and pigeons. The practice of using inbred strains eliminates as far as is possible the influence of individual differences in the subject pool, but has led to the neglect of the study of these differences for their own sake.

Now Louis D. Matzel and collaborators, in what in the future may be seen as a seminal study, have reported the tentative identification of a general learning ability, varying from one individual to another, in mice (*J. Neurosci.* **23**, 6423–6433; 2003). Their work invites independent replication, not least for its 'crossover' character: it remains a rarity for research to combine the skills of the laboratory experimental psychologist with the correlational methods that dominate the study of individual differences.

There are two major elements in the design of the study that should be noted. First of all, Matzel *et al.* used animals from an outbred strain of laboratory mice, on the basis that they show greater behavioural variability — the very opposite of the usual requirement for minimum inter-individual variation (although we have no indication of how laboratory strains might differ from wild populations). Second, five tasks were chosen that differ greatly in their sensory, motor, motivational and information-processing demands, and in which the possibility of transfer of learning from one task to another is minimal.

For instance, the passive avoidance of sensory overload and the active running of a Lashley maze to gain a food reward require quite different information processing and involve different control modalities (aversive as compared with appetitive), different motor activity, and so forth. The other learning tasks were odour discrimination, fear conditioning and a water maze. The animals showed marked variation from individual to individual in all the tasks. The authors also collected some additional measures, including three aspects of free-field behaviour (such as speed of running), the number of faecal pellets deposited during free-field activity, and body weight.

Of the five tasks, the highest correlation

found, 0.47, was between performance in the Lashley maze and on the passive avoidance task. To a physical scientist used to seeing near-exact correspondence, a correlation of 0.47 is a paltry thing. But to a psychologist looking at two such different kinds of task, it is the kind of result that raises the hairs on the back of your neck, and sets the pulse, if not racing, then up to the canter.

In fact, the pattern of correlations for individual mice between all five learning tasks is entirely positive, and although not all of the positive correlations are high, they are sufficient to define a single factor in a factor analysis, which loads the five learning tasks plus the proportion of time spent in the more open areas of the free field. Body weight and number of faecal pellets have no systematic relationships to the other variables, and overall free-field activity and speed of running, although highly correlated with each other, have unsystematic associations with the other variables.

So, have we barked our collective scientific shins against a rodent version of IQ? And do we have here an insight into the nature and origins of human intelligence? The 'ability to learn' is a relatively uncontroversial member of the long list of proposed definitions for intelligence, and Matzel *et al.* have provided some plausible evidence — from impressively controlled experiments — of an influence on learning that cannot be explained by common elements in the tasks. When it comes to the size of the factor, this study yields estimates in the same broad range as has been suggested for the influence of *g*, the general-intelligence factor in humans.

But the sample size, at 56, is a shade uncomfortably low for a factor-analysis study of ten variables: the standard errors of the correlations are of the order of 0.14, so some care is needed in weighing the existence of pattern among the correlations, and in tolerating the fact that some of them are not significantly different from zero, in a conventional bivariate significance test. (The authors kindly made their raw data available to me for independent checks, to ensure that their results are not the outcome of particular choices of factor-analysis methods.)

This is a dilemma for any researcher stepping over the venerable boundary between experimental and correlational approaches to psychology, and Matzel and his collaborators are to be congratulated on striking the right note of caution in reporting their



100 YEARS AGO

A correspondent of the *Times* directs attention to a supposed cure for the mysterious malady known as mountain sickness. The discoverer of the specific is a Russian topographer named Passtoukhof, who, for some years past, has been making ascents in the Caucasus where he has climbed the Grand Ararat, Mount Kasbek, and Mount Elbruz. At such high altitudes as these it is easy to understand that the question of mountain sickness becomes a serious one, and on more than one occasion M. Passtoukhof has found not only himself, but all the other members of his expedition, completely prostrated by it. On one of these occasions it occurred to him to try the experiment of lighting his spirit lamp and making some tea, which he administered to himself and his companions in an almost boiling condition, with a result that far exceeded his expectations. Almost immediately the more serious symptoms disappeared...

From *Nature* 27 August 1903.

50 YEARS AGO

Christianity in an Age of Science. The problem of the relation of religion and science is not dead, as we are often told, but has been made obscure because it has become difficult to define the issue clearly... It is one of the many merits of Prof. Coulson's recent book that he is quite clear on what he is talking about... This little book, consisting of three lectures given under the Riddell Memorial foundation, contains more sound sense on the subject than most works five times its size, for it deals with the ultimate issues and keeps to the point. The challenge of the two universes and two systems of knowledge — the scientific and the religious — is always before his mind. He rejects any theory which would vindicate religion by somehow inserting it into the scientific universe, finding a place for it either in the yet unexplored territory or in the incoherences which can be discovered in scientific conclusions. Perhaps he is a little too hard on theologians who find something of interest to them in Heisenberg's uncertainty principle, for if determinism breaks down in one part of Nature, it may well be only an appearance, or a useful fiction, elsewhere; but undoubtedly he is splendidly right when he says, "If God is here at all, it must be at the beginning of science and right through it".

From *Nature* 29 August 1953.

findings. Spearman's announcement of his discovery of the human general-intelligence factor in 1904 was far less cautious in tone, although from a modern perspective his sampling was less controlled, his sample sizes were of the same order of magnitude, and his experimental controls were, by any standards, rather poor.

A promising pilot study, then? Rather more than that. First of all, we now have five carefully described tasks that can serve to define a reference paradigm, and an example of care in experimental design that should set a benchmark for further work. Second — and this is not immediately obvious from the published paper — there are suggestive pat-

terns buried in the correlation matrix that could be rather productive in determining the next steps, in particular with regard to the relationship between learning and emotionality. And finally, the researchers have performed a signal service in demonstrating how the distinctive skills of what Lee Cronbach once called “the two disciplines of scientific psychology” can combine to open lines of enquiry, and to shed new and refreshing light on the underpinnings of differences in behaviour. ■

Steve Blinkhorn is at Psychometric Research and Development Limited, Brewmaster House, The Maltings, St Albans, Hertfordshire AL1 3HT, UK. e-mail: steve@prd.co.uk

Mechanics

Friction in a spin

Thomas C. Halsey

The mechanics of friction may seem the stuff of high-school physics, but only now are some aspects of the problem being understood. A spinning coin is the subject of a new exploration of frictional forces.

One of the most unlikely revivals of our time is that of the science of mechanics. The first great era of classical mechanics culminated in the work of Joseph-Louis Lagrange (1736–1813) and William Hamilton (1788–1856). But in recent decades, work on nonlinear dynamics, chaos theory, fluid dynamics and the mechanics of biomolecules has driven a highly visible revival of this science. The latest instalment, by Farkas *et al.*¹ in *Physical Review Letters*, on the dynamics of spinning disks, illustrates a convergence of the scientific aesthetic and method that is taking mechanics in a new direction.

The reader can easily study the problem addressed by Farkas and colleagues. Take a coin and launch it across your desk, recording the distance travelled. Now spin the coin about the axis perpendicular to its surface as you launch it (Fig. 1) — the coin will travel farther, even if launched with the same initial velocity. Furthermore, the coin will stop moving and stop spinning at exactly the same instant. Why?

The answer lies in the simplest laws of friction. These laws, known as Amontons' laws, state that the frictional force between two bodies in relative motion is proportional to the force between the bodies (in this case, the weight of the coin) and is directed opposite to the relative motion of the bodies². So the non-spinning coin feels a frictional force opposing its motion, and the magnitude of the force is independent of its velocity. This force will eventually bring the coin to a halt.

Consider now the spinning coin. At one edge the spinning motion increases the relative motion of the coin and surface, and

at the other edge it decreases the relative motion of the coin and the surface (Fig. 1). If the coin spins fast enough, the edge with decreased relative motion will actually move backwards with respect to the surface, even if the overall motion of the coin is forwards. Thus the frictional force from the surface on this edge actually acts to accelerate the coin, causing it to move farther before stopping.

Although striking, this observation is somewhat trivial. The next conclusion drawn by Farkas *et al.*¹ is far less so. The coupled equations describing how the spinning motion of the coin and its velocity both decrease with time are highly nonlinear. Nevertheless, these authors' delicate mathematical analysis proves that not only does the

spinning coin travel farther, but its spinning and its lateral motion stop at the same instant.

Farkas *et al.* also point out another surprising feature of this system. Suppose that we perform the same experiment, not with a coin (whose thickness is much smaller than its diameter) but with a cylinder whose height is comparable to its diameter. The varying frictional forces at the base of the cylinder as it moves will generate a torque (an angular force) about the centre of mass of the cylinder, which will make it bear down more heavily on some parts of its contact footprint, and rest more lightly on other parts. The result is similar to what is known as a Magnus force. Farkas *et al.* show that, as the cylinder moves forwards, its path will tend to curve, much like that of a charged particle in a magnetic field or a fluid (or superfluid) vortex in a steady flow. The direction of path curvature for a charged particle is determined by the sign of its charge, but for the cylinder (as for the vortex) it is determined by the sense of its rotation.

What does all this have to do with the revival of mechanics? No doubt string theorists and creators of Bose–Einstein condensates will be bemused to discover that they are sharing academic departments with colleagues whose idea of fundamental physics involves spinning coins. All give lip service to the notion that exceedingly simple systems can behave in very complex ways, but many are still surprised when faced with a new example of this phenomenon.

Except that a coin sliding on a table is not a simple system. Frictional forces introduce fundamental nonlinearities into the behaviour of mechanical systems, nonlinearities that yield the rich behaviour analysed by Farkas and colleagues. The origin of these nonlinearities lies in a separation of scales. The weight of a coin is not sufficient to distort its shape — a coin is nearly rigid under terrestrial gravity. However, a coin resting on a surface is held up by a relatively small

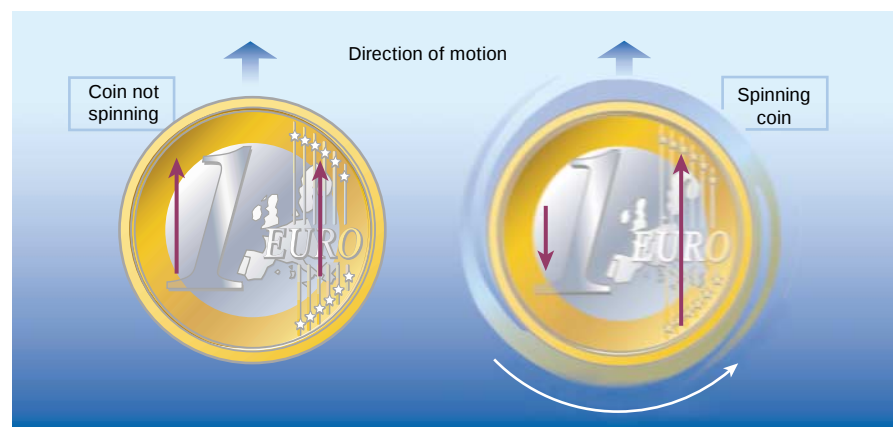


Figure 1 Putting a spin on it. If a coin is projected forwards without spinning (left), each part of the coin has the same velocity relative to the underlying surface (red arrows). If the coin is set spinning as it is projected (right), different parts of the coin have different relative velocities with respect to the surface: one edge of the coin is actually moving backwards relative to the surface and as a result the coin feels a lower net frictional force.

number of micrometre-scale asperities, rough protrusions from the coin's surface. The weight of the coin, channelled through these tiny asperities, creates localized stresses so strong that the metal in the coin flows plastically at the asperities². Thus the frictional mechanics of the coin is actually controlled by material phenomena at scales that are much smaller than the size of the coin. Amontons' laws ignore the existence of these complex phenomena at smaller length scales entirely, at the cost of introducing a highly nonlinear description at the macroscopic scale of the coin.

Granular physics, the science of assemblies of hard grains such as are found in a sand dune or an aspirin bottle, is complex and nonlinear for much the same reason³. The typical forces generated in granular flow and packings are not sufficiently strong to distort the particles macroscopically, so the particles are viewed as rigid. However, on the much smaller scale of the contacts between particles

they can be highly distorted. This separation of scales again yields a highly nonlinear macroscopic description; indeed, there is still considerable controversy over precisely what macroscopic description appropriately encodes the microscopic physics.

Thus the science of mechanics becomes increasingly entwined with asymptotic analysis, the branch of applied mathematics dedicated to understanding the behaviour of equations at the limits where parameters become extremely small or large. Mechanics has moved to centre stage in our endeavour to understand in detail the links between microscopic and macroscopic dynamics. ■

Thomas C. Halsey is at ExxonMobil Research and Engineering, Annandale, New Jersey 08801, USA.
e-mail: thomas.c.halsey@exxonmobil.com

1. Farkas, Z., Bartels, G., Unger, T. & Wolf, D. E. *Phys. Rev. Lett.* **90**, 248302 (2003).
2. Bowden, F. P. & Tabor, D. *The Friction and Lubrication of Solids* (Oxford Univ. Press, Oxford, 2001).
3. Halsey, T. C. & Mehta, A. (eds) *Challenges in Granular Physics* (World Scientific, Singapore, 2003).

Ecology

Tail of death and resurrection

John Harte

Estimating the proportion of rare species in particular habitats is a big issue for ecologists. Hence the intensity of debate over whether 'neutral theory' has predictive value for species abundances.

First, the news. Earlier this year, Brian McGill¹ made an attempt on the life of the 'neutral theory' in ecology. That attempt has been thwarted by Volkov *et al.* (page 1035 of this issue²). But what is the neutral theory? How did the assassination attempt take place? And how has the theory been rescued from that attempt?

The neutral theory is the brainchild of Stephen Hubbell, one of the authors of the new paper². It is a bold attempt to understand the influence of speciation, migration, birth, death, dispersal and extinction on the

composition of ecosystems³. Its boldness lies not in the breadth of processes that it incorporates but in its manifestly oversimplified central assumption of neutrality. This assumption states that differences between traits that we really know differ from species to species — birth, growth and death rates, habitat preferences, dispersal abilities — can be neglected. In the theory, individuals in all species have the same probabilities of birth and death, but random outcomes of probabilistic demographic processes generate differences in species characteristics such as

their abundance. Surely such a theory should be easy to slay?

The attempt on its life was made by McGill in a paper entitled "A test of the unified neutral theory of biodiversity"¹. McGill analysed a data set from Barro Colorado Island (BCI; Fig. 1) in Panama, which contains the location and species identity of all the trees on a 50-hectare plot of tropical forest. The neutral theory predicts a measure known as relative species abundance, which is expressed in terms of a probability distribution describing the fraction of species with any given abundance. The left tail of the probability distribution, which characterizes how many rare species there are, is of particular interest in ecology because it is rare species that are most likely to disappear under habitat loss. It is difficult to include rare species in census counts, so having a theory that reliably predicts how many there are is of enormous value.

McGill contrasted the neutral theory prediction of the relative species abundance with the lognormal distribution, an old standby function often used by ecologists to describe relative species abundances, and concluded that the lognormal fitted the BCI data better than the neutral theory prediction did. A hand-waving justification for the lognormal distribution is that it could arise from the central limit theorem, which in its more familiar form states that the distribution of a variable that is the sum of a collection of random variates is a normal (gaussian) function. If birth and death rates are governed by products of random factors, with a different product chosen for each species, then that same theorem states that those rates will be distributed lognormally from species to species. In contrast to the neutral theory, species differences will be built in from the outset. For many ecologists, the simple plausibility of that argument contrasts with what they feel is the unpalatable central assumption of neutrality, and so McGill's finding that the lognormal distribution fitted the data better was welcome news for many.

To the rescue now come Volkov and two other theoretical physicists, who have teamed up with Hubbell to perform a mathematical *tour de force* and a reanalysis of the data². To calculate the predicted relative species abundance, McGill was forced to perform a computer simulation of the neutral theory, but the complexity of the theory resulted in numerical output whose accuracy could not be quantified. Instead, Volkov *et al.* derive an analytical solution of the neutral theory. When they compare that solution for relative species abundance against the lognormal distribution, the former fits the BCI data somewhat better. So reports of the death of the neutral theory were premature.

And now, the views. A look at Fig. 1 of Volkov and colleagues' paper, reproduced as



Figure 1 Ecological test site: Barro Colorado Island, Panama, on which a 50-hectare Forest Dynamics Plot is maintained by the Center for Tropical Forest Science of the Smithsonian Tropical Research Institute. Tree census data have been gathered here since 1982; it is these data that are used in the tests^{1,2} of neutral theory discussed here.

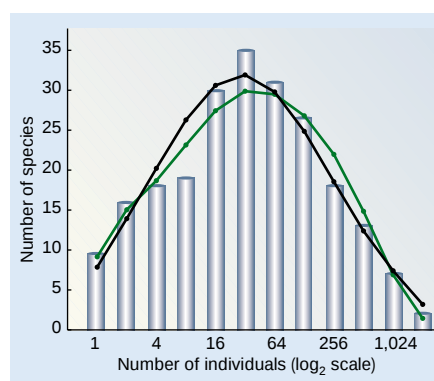


Figure 2 Matching theory to data. Lognormal (black) and neutral (green) expectations against the data on 21,457 trees in 225 species from the 50-hectare plot on Barro Colorado Island. Here the data are grouped in 12 logarithmic intervals; note the slight asymmetry in the neutral theory prediction, favouring the bulged left-hand tail of the empirical distribution. Reproduced from ref. 2.

Fig. 2 here, might leave readers perplexed. Both the lognormal and the neutral theory predictions look very good and at the tails of the distribution they are essentially indistinguishable. Surely when 21,457 individual trees, distributed among 225 tree species, are counted there will be measurement uncertainty, particularly given that very tiny young trees are not included. Should we really care whether one function fits slightly better than the other? More generally, how should theory-testing in ecology proceed?

Proposed theories are valuable when they make many falsifiable predictions, preferably about a variety of phenomena not previously recognized as being interconnected. Arguments about whether this or that function fits an empirical relative species abundance slightly better are unlikely to advance the field. Consider the lognormal distribution first. It is not really a theory, but rather a proposed mathematical function; its connection to relative species abundance has been motivated by a conceptual model of how growth and death are regulated in ecology. Advocates of the lognormal distribution would best serve their cause if they actually examined and modelled the dominant mechanisms of growth and death, checked for the applicability of the central limit theorem, and then made a swarm of testable predictions — not just about relative species abundance but about growth and death rates under different circumstances, about the relationship between body size and abundance, about the spatial distributions of species, and about time series of population fluctuations under different environmental conditions.

The ability of the neutral theory to make testable predictions about a wide variety of behaviours could also be more fully

explored. One prediction of the theory pertains to what ecologists call beta diversity — the pattern of changing species composition in separated patches of habitat as a function of the distance between patches. In the neutral theory, species turnover with distance should result from dispersal but not from spatial variability in the different habitat niches. A test of this at the BCI plot found empirical patterns to be inconsistent with the prediction of the neutral theory⁴.

Indeed, theories such as Hubbell's are valuable because they can and will fail some tests. In failing, they will tell us about the importance of the mechanisms that they assume away at the outset. Perhaps ecologists will some day develop a theory whose simplifications resemble that of the ideal

Quantum gravity

An astrophysical constraint

Sean Carroll

A quantum theory of gravity is proving elusive. Observations of radiation from the Crab nebula now place even stronger constraints on the likelihood of detecting the effects of quantum gravity.

For decades, physicists have been in search of quantum gravity, a theory that would encompass both general relativity and quantum mechanics. One of the difficulties in this quest is the paucity of detailed experimental data; direct effects of quantum gravity are generally thought to be out of reach of foreseeable experiments. But, if we are lucky, the effects of quantum gravity might cause small violations of spacetime symmetries, violations that could be directly observed through the behaviour of particles at high energies or over large distances. A new astrophysical test — reported by Jacobson *et al.*¹ on page 1019 of this issue — suggests, however, that we might not get lucky.

In the process of developing his theory of blackbody radiation in the late nineteenth century, Max Planck was forced to introduce a new fundamental constant into physics — what we now know as Planck's constant, h . Almost immediately, Planck realized that this number could be combined with the speed of light, c , and Newton's gravitational constant, G , to construct a set of natural units for quantities such as length, energy, mass and time (known as the Planck length, the Planck energy, and so on). He was happy to note that even extraterrestrials would understand these units, thus making quantitative communication between civilizations possible.

Planck's units combine the fundamental parameters of special relativity (c), general relativity (G) and quantum mechanics (h). We therefore expect that they should indi-

gas assumption in thermodynamics and statistical physics: a seemingly preposterous assumption (point molecules, purely elastic collisions) that yields amazingly accurate predictions of a multitude of phenomena. No physicist would say that, because PV does not exactly equal nRT , the theory is wrong. Let's hope that ecologists heed that message.

John Harte is in the Energy and Resources Group, University of California, Berkeley, California 94720, USA.

e-mail: jharte@socrates.berkeley.edu

1. McGill, B. J. *Nature* **422**, 881–885 (2003).
2. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. *Nature* **424**, 1035–1037 (2003).
3. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
4. Condit, R. *et al.* *Science* **295**, 666–669 (2002).

cate the point at which quantum gravity begins to manifest itself. For example, if two particles collide with a total energy greater than the Planck energy, quantum gravity should have an important role in the outcome. Unfortunately, the Planck energy scale is approximately 10^{19} gigaelectronvolts — 16 orders of magnitude larger than the regimes being explored by our highest-energy particle accelerators. So it seems difficult, even impossible, to get any direct experimental data relevant to the reconciliation of gravity and quantum mechanics.

But our understanding of fundamental physics will not be coherent, much less correct, until quantum gravity is understood. It is therefore worth trying to think of clever ways to access phenomena at the Planck scale that we might at first think are beyond our reach. One possibility, common to several (but certainly not all) models of quantum gravity, is a breakdown of Lorentz invariance. This symmetry, a cornerstone of special relativity, says that there is no universal standard of rest; reference frames that are moving with respect to each other at constant velocities should be physically equivalent, and the speed of light looks the same to any observer. But perhaps this is only an approximation; if spacetime is discrete (rather than smooth) at small scales, for example, Lorentz invariance may break down at very short distances or very high energies.

Astrophysical phenomena provide a way of testing this idea. A phenomenological model² has been proposed in which Lorentz-



Figure 1 The Crab nebula. Located in the constellation of Taurus, this cloud of gas is all that remains of a star whose explosion in AD 1054 was recorded by Chinese astronomers. In this composite image (from the Chandra observatory and the Hubble Space Telescope), the blue regions are synchrotron radiation, generated by high-energy electrons whirling in the system's magnetic field. Jacobson *et al.*¹ have analysed the frequency spectrum of this radiation for a signature of quantum gravity — the theory uniting quantum mechanics and relativity. They find none.

violating effects are introduced into the so-called dispersion relations that relate the energy and momentum of particles. As a result, photons of different frequencies should travel at slightly different velocities; such an effect could potentially deform pulses of radiation from distant γ -ray bursts in an observable way. This approach introduces free parameters that can be expressed as dimensionless numbers divided by the Planck energy. Thus, if these numbers can be constrained to magnitudes less than unity, this sort of hypothetical quantum-gravity effect can be ruled out even at the Planck scale.

If we limit ourselves to considering photons and electrons, we seek to constrain two dimensionless numbers parameterizing Lorentz violation. Perhaps surprisingly, results from a variety of astrophysical and laboratory tests have already excluded a wide range of possible values for these parameters, to levels beyond the Planck scale; nevertheless, substantial regions have remained

open. But now Jacobson *et al.*¹ are able to essentially rule out any Lorentz violation of this type at Planck energies or below, at least under certain dynamical assumptions. (Almost simultaneously, laboratory experiments³ have placed constraints that are not

quite as strong but are also beyond the Planck scale.)

Jacobson *et al.* consider synchrotron radiation, emitted by electrons circling in a magnetic field, from the Crab nebula (Fig. 1). To produce high-energy photons through synchrotron emission, the electrons must be moving close to the speed of light. If Lorentz invariance is violated, the maximum velocity for photons and for electrons can have a slightly different value, which imposes a cut-off on the frequency of synchrotron radiation that can be produced. Using observations of radiation at frequencies beyond this cut-off, Jacobson *et al.* are able to set the new stringent limit on Lorentz invariance. The crucial assumption made in their analysis is that the behaviour of photons and electrons can be described by an 'effective local field theory' at low energies. Such theories are well used in this area of physics, and this seems a reasonable assumption to make, but exceptions are known. So a window, albeit small, remains open for Planck-scale effects.

Searching for effects of quantum gravity through violations of Lorentz invariance is an optimist's game; it is certainly plausible that such violations are strictly absent from the correct theory. But optimism is necessary if we are to make progress in this field. Jacobson and colleagues' results¹ apply to perhaps the most optimistic proposal for observable Lorentz violation, in which new effects are suppressed by only a single power of the Planck scale; another realistic possibility is that Lorentz symmetry is indeed violated, but only at quadratic order. We should therefore remain hopeful, and search even harder for direct ways to detect increasingly subtle effects of spacetime structure on very small scales. ■

Sean Carroll is at the Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA.

e-mail: carroll@theory.uchicago.edu

1. Jacobson, T., Liberati, S. & Mattingly, D. *Nature* **424**, 1019–1021 (2003).
2. Amelino-Camelia, G., Ellis, J., Mavromatos, N. E., Nanopoulos, D. V. & Sarkar, S. *Nature* **393**, 763–765 (1998).
3. Myers, R. C. & Pospelov, M. *Phys. Rev. Lett.* **90**, 211601 (2003).

Mars

The devil is in the dust

Conway B. Leovy

Mars is a highly dynamic planet — at least as far as dust is concerned. A better knowledge of how dust is lofted into the atmosphere will help to untangle the complex evolutionary history of the planet's surface.

New laboratory work by Greeley *et al.*¹ and numerical flow simulations by Toigo *et al.*², both appearing in the *Journal of Geophysical Research*, provide insights into the role of 'dust devils' in injecting dust into the martian atmosphere.

Dust transport is a key to the long-term evolution of the martian surface, much of which has undergone episodes of erosion, deposition, burial and exhumation over the past four billion years³. Although flowing water is believed to be responsible for some

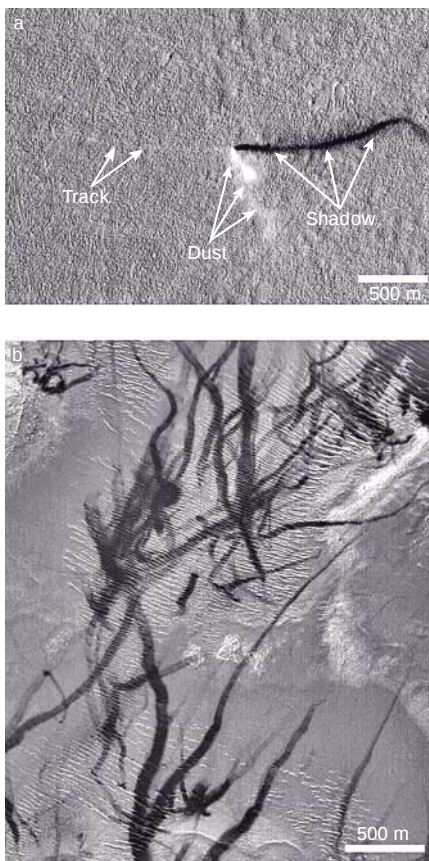


Figure 1 Evidence of dust devils: exhibits a and b. a, In this picture, taken on 10 April 2001, the Sun is illuminating the surface of Mars from the left, so creating a dark shadow of the wispy dust devil which can be seen in the centre of the image. The faint line to the left is the track of the dust devil; its height, estimated from the shadow, was just over 1 km. b, Streaks on the martian surface, interpreted as dust-devil tracks. Such tracks often run across surface features such as dunes and boulders. They are usually dark (but not always, as in a), which is thought to be due to removal of bright dust, revealing darker, underlying material. The area depicted here is about 3 km by 5 km. Both images were taken by the Mars Orbiter Camera.

of this surface change, lifting and redistribution of dust by wind has had an important role over that entire period, and may be the largest factor today. Suspended dust also influences atmospheric dynamics by absorbing solar radiation, efficiently heating the thin atmosphere and thereby contributing to the driving of large-scale winds. Atmospheric dust content has a seasonal maximum when Mars is closest to the Sun, but the maximum dust load varies from one seasonal cycle to the next. Occasionally, large dust storms develop in several areas simultaneously, spreading a dense pall over the planet and obscuring much of the surface for weeks at a time⁴.

Computer models of Mars' atmospheric general circulation do a reasonable job of simulating the dynamics of the dust-laden

atmosphere^{5,6}, but they are constrained by our limited understanding of how dust is lifted from the surface. The stress exerted on the surface by wind can lift dust particles into suspension. Because of the low atmospheric density, the minimum threshold wind speed for lifting is an order of magnitude larger on Mars than on Earth. Moreover, dust particles (of diameter less than 60 μm) require much stronger winds for lifting than do small, sand-size particles (roughly 60–200 μm), because inter-particle cohesion effectively locks the smaller dust particles into the surface. Dust is injected into the atmosphere when sand-size particles are lifted and hop or 'saltate' along the ground, knocking the dust particles into suspension⁷. This mechanism requires the coexistence of dust and sand-size particles, and martian surface wind speeds in excess of 30 m s^{-1} . Such high wind speeds are relatively rare, and sand-size particles may not be very widespread. Nevertheless, large dust storms are remarkably common. More than 5,000 have been observed during two martian annual cycles in images obtained with the Mars Orbiter Camera on board the Mars Global Surveyor orbiter⁸.

Injection of dust into the martian atmosphere by a distinctly different phenomenon — dust devils — has also been widely observed^{3,9,10}. Dust devils are common in arid regions on Earth during the hot season: strong daytime heating of the surface generates intense turbulence and convective plumes, and air converging into these plumes tends to conserve any initial angular momentum. If the plume and initial angular momentum are sufficiently intense, a spinning vortex develops that picks up dust to form the visible dust devil (Fig. 1a). Because of the intense daytime heating, such convectively driven 'boundary-layer vortices' on Mars can be much larger than on Earth, with diameters up to 1 km and heights up to 8 km. Surface streaks produced by dust devils are widespread³ (Fig. 1b). Most of these apparently form when the dust devil removes a thin layer of bright dust from a darker substrate. Subsequent sedimentation of dust can rapidly cover these streaks.

It has long been suspected that dust is more easily lifted by dust devils than by wind stress and saltation alone⁷. The new experiments verify this and demonstrate the lifting mechanism. Greeley *et al.*¹ have generated laboratory vortices whose scaled pressure and flow fields closely match those observed in terrestrial dust devils and in vortices observed at the Mars Pathfinder landing site. They investigated thresholds for particle lifting for a wide range of particle diameters and densities, and a range of vortex diameters, at pressures ranging from terrestrial to martian. In addition to lifting by the wind-stress mechanism, low pressure in the laboratory vortex core induces anomalous upward

pressure on particles that counters inter-particle cohesion. This 'vacuum-cleaner effect' acts most efficiently on the smallest particles, so that dust is lifted about as easily as sand-size particles.

If low dust-devil core pressure works in the same way when scaled up to martian convective vortex conditions, these experiments show that the wind-speed threshold for dust lifting by convective vortices would be much lower than for wind stress alone, and that saltation by sand-size particles would not be required. This may explain the surprising ubiquity of dust devils and wind streaks on Mars. Whether a similar vacuum-cleaner effect caused by intense small vortices within larger dust storms has a role in lifting dust is not known, but this possibility may help to explain the large number and wide distribution of the large dust storms.

Computer flow-simulation models can also provide insights into the dynamics of boundary-layer vortices. By adapting such a model to martian conditions, Toigo *et al.*² have been able to simulate the evolution of convective vortices with and without strong background wind. These vortices closely resemble the vortices seen on Earth and Mars, as well as the laboratory vortices produced by Greeley and colleagues. They form distinct coherent structures, where intense updraughts enhance rotating motion through convergence of the horizontal flow. Although the simulated Mars vortex wind and pressure anomalies do not yet reach the threshold for dust lifting found by Greeley *et al.*, such computer simulations promise to be an important tool for interpreting laboratory and field observations of dust devils. Integration of the laboratory dust-lifting results with computer vortex models may soon permit reliable simulation of the complete process of dust injection into the atmosphere by dust devils. Together, these new results^{1,2} significantly advance our ability to simulate dust lifting, and promise to contribute to our understanding of martian surface evolution. ■

Conway B. Leovy is Professor Emeritus of Atmospheric Sciences, University of Washington, Seattle, Washington 98105, USA.

e-mail: conway@atmos.washington.edu

- Greeley, R. *et al.* *J. Geophys. Res.* **108**, doi:10.1029/2002JE001987 (2003).
- Toigo, A., Richardson, M., Ewald, S. & Gierasch, P. *J. Geophys. Res.* **108**, doi:10.1029/2002JE002002 (2003).
- Malin, M. & Edgett, K. *J. Geophys. Res.* **106**, 23429–23570 (2001).
- Zurek, R. *et al.* in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 835–933 (Univ. Arizona Press, Tucson, 1992).
- Leovy, C. *Nature* **412**, 245–249 (2001).
- Murphy, J. *et al.* *J. Geophys. Res.* **100**, 26357–26376 (1995).
- Greeley, R., Lancaster, N., Lee, S. & Thomas, P. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 730–766 (Univ. Arizona Press, Tucson, 1992).
- Cantor, B. *6th Int. Mars Conf.* Pasadena, California, Abstr. 3166 (Lunar & Planetary Inst., Houston, Texas, 2003).
- Thomas, P. & Gierasch, P. *Science* **230**, 175–177 (1985).
- Metzger, S., Carr, J., Johnson, J., Parker, T. & Lemmon, M. *Geophys. Res. Lett.* **26**, 2781–2784 (2002).

Biomaterials

Silk's secrets

Edward Atkins

Despite centuries of human use of silk fibres from silkworm cocoons, and an emerging industry devoted to making artificial silk, questions remain about how insects produce it. New work *in vitro* tackles the problem.

Once solely the purview of insects and spiders, the production of silk fibres is now a biotechnological reality. Fibrous silk is used widely, in materials from clothing to carpets to parachutes, so there's a great deal of interest in understanding the precise details of how it forms from silk proteins — whether *in vivo* or in artificial circumstances. Hence the importance of the paper on page 1057 of this issue, where Jin and Kaplan¹ describe several early steps in fibre production.

The traditional method of obtaining silk involves meticulously unravelling the fibres from silkworm cocoons (Fig. 1) and weaving them into fabrics. But other sources of silk fibres are now available. For example, Nexia Biotechnologies, based in Montreal, has developed 'transgenic' goats that express spiders' silk proteins in their milk². These proteins have been processed into fibres with the registered trade name BioSteel.

The production, or spinning, of silk fibres by silkworms and spiders incorporates many facets of polymer physics, as well as chemical and biochemical control³. Properties such as strength, toughness and (in certain spider silk fibres, for instance) elasticity are now being understood at a fundamental level and can be manipulated. For example, the dragline silk of the spider *Nephila clavipes* — hailed as a 'super-fibre'^{2–5} — is much stronger and more extensible than traditional silk from the silkworm *Bombyx mori*. Indeed, dragline spider silk has been reported to have a strength only 15% lower than that of high-tensile engineering steel, and, because of the much lower relative density of silk (17% of that of steel), it is a far stronger material³.

In terms of manipulation, Shao and Vollrath⁶ have shown that, by changing the reeling conditions, silkworm silks can be made stronger, stiffer and more extensible, approaching the properties of spider dragline silk. And the opportunity exists for modified silk proteins — designed to give even better silk fibres — to be genetically engineered and produced in large quantities, thereafter being mechanically spun using controlled water content and elongational stress. This would form the basis for a renaissance in silk materials. Given these many possibilities, it is important that any remaining puzzles about the mechanism of silk spinning should be solved.

Silk proteins — known as silk fibroins — are stored in the glands of insects and spiders as an aqueous solution. During the spinning process, by which fibres are produced from this silk 'dope', the concentration of silk in the solution is gradually increased, and finally elongational stress is applied to produce a partly crystalline, insoluble fibrous thread in which the bulk of the polymer chains in the crystalline regions are oriented parallel to the fibre axis.

It is known that silk fibroin consists of both hydrophilic and hydrophobic regions — it is a block-like polymeric system. But what scientists are still not sure about is how this complex silk protein can be maintained in a concentrated silk dope without fear of irreversible precipitation or crystallization, potentially blocking the whole spinning device. It is understood that the water is acting as a plasticizing agent, keeping the protein malleable, but the nature of the silk itself in this environment remains largely

unknown. Nor is it known how the protein changes in structure, texture or morphology with increasing concentration, finally emerging, after the final stages of elongational stress, as an insoluble fibre.

Enter Jin and Kaplan¹, who, in their elegant experiments, have followed the behaviour of silk solutions as a function of a controlled decrease in the water content. Starting with cocoons of *B. mori*, the authors first 'degummed' them — removing the sericin proteins, which glue the fibroins together in the cocoon — and then redissolved the silk fibroin. Next they created an osmotic stress by adding gradually increasing amounts of the high-molecular-mass compound polyethylene glycol (either mixed in with the fibroin dope or segregated by dialysis membranes), which preferentially competes with the silk fibroin for the water molecules in which the protein is dissolved.

The results show the development of phase separation and the appearance of nanoscale colloidal-like particles (micelles) as water levels are reduced. These micelles are believed to form by the folding of silk chains through their hydrophilic blocks, thereby allowing the hydrophobic blocks to be in close proximity; the folded entities then associate with each other to generate spherical micelles with diameters in the range 100–200 nm. The micelles themselves aggregate as the water content is decreased still further, forming micrometre-scale globules and gel-like states.

The appearance of similar globular structures in the fracture surfaces of native silk fibres supports the contention that these laboratory experiments are mimicking the natural process. The authors suggest that an explanation of this phase behaviour lies in the hydrophobic/hydrophilic partitioning and folding of the silk fibroin — both of which are processes regulated by the water content. Jin and Kaplan show further that adding directional stress causes deformation of the globules, which then line up in the same orientation.

Quite how the deformed globules then form fibres is a question for the future. But Jin and Kaplan's findings should already prove useful, providing a basis for researchers to investigate how far it is possible to manipulate silk spinning, and how to engineer new silk proteins that can still form fibres.

Edward Atkins is in the Department of Physics, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK.

e-mail: e.atkins@bristol.ac.uk

1. Jin, H.-J. & Kaplan, D. L. *Nature* **424**, 1057–1061 (2003).
2. Lazaris, A. et al. *Science* **295**, 472–476 (2002).
3. Vollrath, F. & Knight, D. P. *Nature* **410**, 541–548 (2001).
4. Kaplan, D. L. et al. *Silk Polymers: Materials Science and Biotechnology* (Am. Chem. Soc., Washington, 1994).
5. Vollrath, F., Madsen, B. & Shao, Z. *Proc. R. Soc. Lond. B* **268**, 2339–2346 (2001).
6. Shao, Z. & Vollrath, F. *Nature* **418**, 741 (2002).



Figure 1 Silk maker: *Bombyx mori*.

High-energy astronomy

Flashes from planet smashes

Astrophys. J. (in the press); Preprint
http://arxiv.org/abs/astro-ph/0308218 (2003)

Colliding planets around other stars produce flashes of light that should be detectable from Earth, say Bing Zhang and Steinn Sigurdsson. They estimate that there should be about 50 such events per year in the Milky Way galaxy, at least some of which would be close enough to be seen by ultraviolet or X-ray observatories.

The researchers calculate that a head-on smash between a planet of Jupiter's size and one like Earth would produce a flash lasting for several hours, with a peak wavelength in the extreme ultraviolet or soft X-ray part of the spectrum. This should be detectable by satellites such as the Extreme Ultraviolet Explorer (EUVE), which operated for eight years in the 1990s, or current X-ray satellites such as Chandra and XMM-Newton. Zhang and Sigurdsson think that up to ten such events might be found in the EUVE archives, and several in the data from the X-ray satellites.

The giant planet survives such a collision but continues to glow in the infrared for thousands of years. Bodies like this might be visible in nearby star-forming regions as 'hot' regions distinct from their parent stars.

Philip Ball

Evolutionary biology

Golgi traces revealed

Proc. R. Soc. Lond. B doi: 10.1098/rsbl.2003.0058 (2003)

Joel B. Dacks *et al.* have challenged the view that a handful of eukaryotic cell lineages lack Golgi bodies. These membrane-bounded compartments move proteins around inside cells, and are thought to have contributed to the runaway success of eukaryotes.

Golgi bodies are virtually omnipresent in eukaryotes (loosely speaking, those organisms whose cells have defined nuclei). 'Golgi-lacking' organisms were traditionally thought to have branched off from other eukaryotes before the organelles were acquired, some 1–2 billion years ago.

By combing the genetic sequences of five Golgi-lacking organisms, however, Dacks *et al.* found evidence of genes similar to those that make Golgi components in mammals and yeast. The creatures represent four of the seven proposed Golgi-lacking eukaryotic lineages, and the findings include the first evidence for such genes in two of these, the heteroloboseids and pelobionts.

The authors argue that the organisms might once have possessed the visible stack of compartments typical of a Golgi body, but that the Golgi genes might now be involved in forming a differently shaped

Golgi that traffics proteins via small vesicles. They conclude that the last common ancestor of eukaryotes did possess a Golgi body, but it is not known exactly how it was acquired.

Helen Pearson

Neuroscience

How to forget

Neuron **39**, 655–669 (2003)

Some memories are with us for minutes, whereas others endure for years. Certain proteins help to move short-term memories into long-term storage, but, in a bid to prevent us from remembering absolutely everything, their effects seem to be balanced by a second group of molecules — memory suppressors.

One family of mammalian proteins, the C/EBPs, contains representatives from both camps. In an attempt to understand their function, Amy Chen *et al.* have used a novel protein-targeting approach to damp down the influence of the memory-suppressing C/EBP proteins in the brains of live mice, leaving the memory-storing members of the family unaffected. The authors found that these animals acquired long-term memories more easily. In a tricky test situation, the modified mice remembered the location of a hidden platform better than controls.

The results show how altering the balance of C/EBP proteins can modify the conversion of short-term to long-term memories. As a family, the proteins are known to act as gene-transcription factors, so the authors suggest that they might affect the expression of other memory-related genes.

Helen R. Pilcher

Ecology

Under attack

Ecol. Lett. **6**, 712–715 (2003)

According to Anurag A. Agrawal and Peter M. Kotanen, herbivores attack foreign plant species just as much as native plants. This runs counter to the idea that invasive species are successful because they escape the natural enemies present in their homelands.

Agrawal and Kotanen conducted a field experiment in Ontario, Canada, using 30 plant species divided into 15 pairs of closely

related American and Eurasian plants (see pictures). They found that the Eurasian species sustained significantly more damage from herbivores. The authors suggest that native plants might be better adapted to resist the local herbivores. Similarly, a laboratory test found that exotic species were just as palatable as their native counterparts to a caterpillar.

The finding challenges the 'enemy release' hypothesis that exotic species in general face lower levels of attack. Instead, Agrawal and Kotanen suggest, the species most likely to escape enemies might be those lacking close relatives — and pre-adapted herbivores — in the native flora.

John Whitfield

Developmental biology

Creating diversity

Dev. Cell **5**, 231–243 (2003)

There are many different cell types in a multicellular organism, yet all of them originated from the same founder cell. Diversity is created through asymmetric cell divisions — divisions that produce not two identical daughter cells, but two different ones. Kate M. O'Connor-Giles and James B. Skeath now propose a new role for the Sanpodo protein during asymmetric cell division in the fly brain.

One way to create different cell fates is to switch on the so-called Notch signalling pathway in one daughter cell (A) and turn it off in the other (B). Selective delivery of the Numb protein to the B cell effectively numbs all Notch signals. Sanpodo, on the other hand, specifies the A (Notch-dependent) fate — even though it is present in A and B cells. The details of both processes are unknown.

O'Connor-Giles and Skeath show that, in A cells, Sanpodo is free to move to the cell surface, where it is found associated with the Notch receptor. But in B cells, Numb binds to Sanpodo and keeps it intracellular. These findings support a model in which the Numb protein blocks Notch signalling by keeping Sanpodo captive inside the cell. This prevents Sanpodo from associating with the Notch receptor at the cell surface and bringing about the A-cell fate.

Marie-Thérèse Heemels



Native and foreign: *Lactuca canadensis* (left) and *L. serriola* (right).

Plasma antioxidants from chocolate

Dark chocolate may offer its consumers health benefits the milk variety cannot match.

There is some speculation that dietary flavonoids from chocolate, in particular (–)epicatechin, may promote cardiovascular health as a result of direct antioxidant effects or through antithrombotic mechanisms^{1–3}. Here we show that consumption of plain, dark chocolate (Fig. 1) results in an increase in both the total antioxidant capacity and the (–)epicatechin content of blood plasma, but that these effects are markedly reduced when the chocolate is consumed with milk or if milk is incorporated as milk chocolate. Our findings indicate that milk may interfere with the absorption of antioxidants from chocolate *in vivo* and may therefore negate the potential health benefits that can be derived from eating moderate amounts of dark chocolate.

To determine the antioxidant content of different chocolate varieties, we took dark chocolate and milk chocolate prepared from the same batch of cocoa beans and defatted them twice with *n*-hexane before extracting them with a mixture of water, acetone and acetic acid (70.0:29.8:0.2 by volume). We measured their *in vitro* total antioxidant capacities using the ferric-reducing antioxidant potential (FRAP) assay⁴; FRAP values were 147.4 ± 4.5 and $78.3 \pm 3.4 \mu\text{mol}$

reduced iron per 100 g for dark and milk chocolate, respectively. Volunteers must therefore consume twice as much milk chocolate as dark chocolate to receive a similar intake of antioxidants.

We recruited 12 healthy volunteers (7 women and 5 men with an average age of 32.2 ± 1.0 years (range, 25–35 years). Subjects were non-smokers, had normal blood lipid levels, were taking no drugs or vitamin supplements, and had an average weight of 65.8 ± 3.1 kg (range, 46.0–86.0 kg) and body-mass index of $21.9 \pm 0.4 \text{ kg m}^{-2}$ (range, 18.6–23.6 kg m^{-2}). On different days, following a crossover experimental design, subjects consumed 100 g dark chocolate, 100 g dark chocolate with 200 ml full-fat milk, or 200 g milk chocolate (containing the equivalent of up to 40 ml milk).

One hour after subjects had ingested the chocolate, or chocolate and milk, we measured the total antioxidant capacity of their plasma by FRAP assay. Plasma antioxidant levels increased significantly after consumption of dark chocolate alone, from $100 \pm 3.5\%$ to $118.4 \pm 3.5\%$ (*t*-test, $P < 0.001$), returning to baseline values ($95.4 \pm 3.6\%$) after 4 h (Fig. 2a). There was no significant change in plasma FRAP values over the same period after ingestion of milk chocolate alone or of dark chocolate with milk (Fig. 2a).

The areas under the curves of (–)epicatechin plasma levels plotted against time⁵ were measured over the same 4-h period after ingestion for the three different conditions. Absorption of (–)epicatechin into the bloodstream after ingestion of chocolate was significantly less when the chocolate was accompanied by milk ($-46.4 \pm 4.1\%$; analysis of variance (ANOVA), $P < 0.001$) or if the chocolate itself contained milk ($-69.1 \pm 3.9\%$; ANOVA, $P < 0.001$; Fig. 2b).

Addition of milk, either during ingestion or in the manufacturing process, therefore inhibits the *in vivo* antioxidant activity of chocolate and the absorption into the bloodstream of (–)epicatechin. This inhibition



Figure 1 Stack of benefits? Unlike its milky counterpart, dark chocolate may provide more than just a treat for the tastebuds.

could be due to the formation of secondary bonds between chocolate flavonoids and milk proteins^{6,7}, which would reduce the biological accessibility of the flavonoids and therefore the chocolate's potential antioxidant properties *in vivo*.

Our findings highlight the possibility that the *in vivo* antioxidant activity of flavonoids could be impaired by other dietary constituents. Other food combinations may also counteract the absorption and protective effects of flavonoids. There is therefore a need to take into account dietary habits when designing studies to assess the association between flavonoid-rich foods, antioxidant activity and degenerative diseases.

Mauro Serafini*, **Rossana Bugianesi***,
Giuseppe Maiani*, **Silvia Valtuena***,
Simone De Santis*, **Alan Crozier†**

*Antioxidant Research Laboratory, Unit of Human Nutrition, National Institute for Food and Nutrition Research, Via Ardeatina 546, 00178 Rome, Italy

e-mail: serafini@inran.it

†Plant Products and Human Nutrition Group, Graham Kerr Building, Division of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

- Rein, D. et al. *J. Nutr.* **130**, 2109S–2114S (2000).
- Holt, R. R., Schramm, D. D., Keen, C. L., Lazarus, S. A. & Schmitz, H. H. *J. Am. Med. Assoc.* **287**, 2212–2213 (2002).
- Steinberg, F. M., Bearden, M. N. & Keen, C. L. *J. Am. Diet. Assoc.* **103**, 215–223 (2003).
- Benzie, I. F. F. & Strain, J. J. *Anal. Biochem.* **239**, 229–240 (1996).
- Maiani, G., Serafini, M., Salucci, M., Azzini, E. & Ferro-Luzzi, A. *J. Chromatogr. B* **692**, 311–317 (1997).
- Charlton, A. J. et al. *J. Agric. Food Chem.* **50**, 1593–1601 (2002).
- Serafini, M., Ghiselli, A. & Ferro-Luzzi, A. *Eur. J. Clin. Nutr.* **50**, 28–32 (1996).

Competing financial interests: declared none.

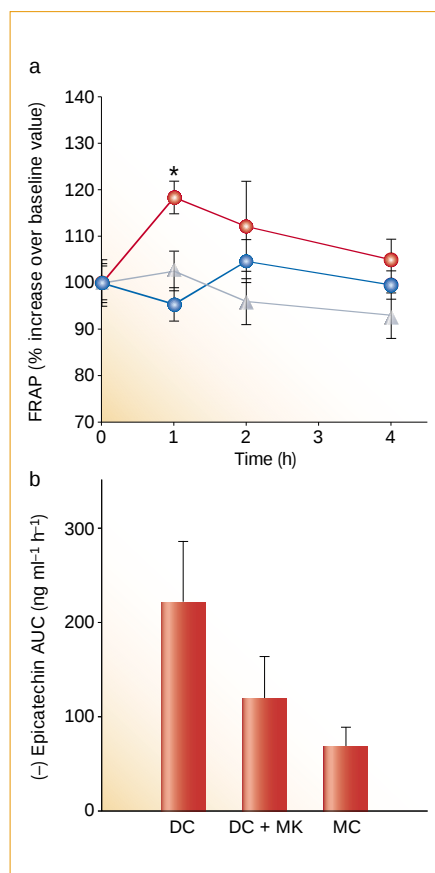


Figure 2 Effects of acute ingestion of 100 g dark chocolate (DC), 100 g dark chocolate with 200 ml milk (DC + MK) or 200 g milk chocolate (MC) on the total antioxidant capacity (TAC) and (–)epicatechin content of human plasma. **a**, Mean TAC of plasma samples at the indicated times after chocolate consumption, expressed as ferric-reducing antioxidant potential (FRAP)⁴. Values are mean percentage increases ($\pm$ s.e.m.) relative to baseline values ($n = 12$). Red circles, DC; blue circles, DC + MK; grey triangles, MC. Asterisk denotes $P < 0.001$. **b**, Mean (–)epicatechin levels in plasma, expressed as the area under the curve in **a** (AUC, in $\text{ng ml}^{-1} \text{h}^{-1}$) for the 4-h period after chocolate consumption. Values are significantly different from one another (see text).

Like-charged particles at liquid interfaces

Nikolaides *et al.*¹ propose that the puzzling attraction that occurs between micrometre-sized particles adsorbed at an aqueous interface is caused by a distortion of the liquid interface that is due to the dipolar electric field of the particles and which induces a capillary attraction. Here we argue that this effect cannot account for the observed attraction, on the fundamental grounds that it is inconsistent with force balance.

To estimate the influence of the surface deformation, the authors assume that the sum total of the electrostatic pressure acting on the liquid interface is equivalent to an external force, F , pushing the particle into the water. The resulting deformation of the interface would give rise to a long-range interparticle interaction energy

$$U(r) = (F^2/2\pi\gamma) \ln(r/r_0) \quad (1)$$

where γ is the surface tension, r is the distance between particles, and r_0 is an arbitrary constant. However, the electrostatic force acts on the particle and the liquid interface simultaneously, so equation (1) does not apply.

The force F on the sphere is balanced by surface tension, creating a dimple in the water surface. The shape of the dimple is governed by the Young–Laplace equation

$$[(1/R_1) + (1/R_2)]\gamma = \Delta p \quad (2)$$

If the force acts only on the particle, then the pressure difference across the interface $\Delta p \approx 0$ and the radii of curvature R_1 and R_2 of the surface are equal but opposite (Fig. 1). The resulting water level around an isolated sphere would be $h(r) = F/(2\pi\gamma) \ln(r/r_0)$ for small surface slopes, such that $F \approx 2\pi r_c \gamma h'(r_c)$, where $2\pi r_c$ is the length of the contact line. Owing to the long range of the logarithm, the force on the sphere is

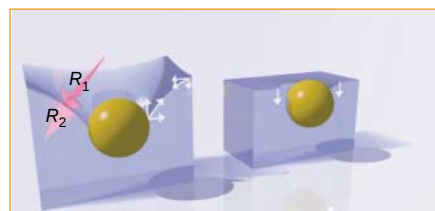


Figure 1 Pushing a sphere into water creates a dimple (left) in the surface, caused by the force on the particle being balanced by the surface tension. But if the electrostatic force pushing the sphere into the liquid is balanced by the electrostatic pressure on the liquid interface, there is no force on the rim and the surface is flattened again (right), inhibiting capillary attraction. R_1 , R_2 are the radii of curvature of the surface.

passed on to the rim of the vessel containing the liquid (Fig. 1, left).

The situation changes fundamentally when the force F is not external but originates from a surface pressure $\Delta p \propto 1/r^6$, as stated in ref. 1. Now the meniscus profile is $h(r) = -F/(8\pi\gamma)(r_c/r)^4$ — that is, very short range. The electrostatic force F pushing the sphere into the liquid is balanced by the sum total of electrostatic pressure on the liquid interface and there is no force on the rim of the vessel (Fig. 1, right). In contrast, by unquestioningly applying equation (1), one implicitly admits that there is also a force on the vessel, a force not balanced by any other force. This is inconsistent with Newton's third law.

Capillary attraction between spheres is caused by the overlap of their dimples, which reduces the total surface area of the water². The resulting attraction energy is

$$U(r) = -(F^2/\pi\gamma)(r_c/r)^6 \quad (3)$$

for large r . This interaction has a much shorter range than that shown in equation (1), even a shorter range than the dipole–dipole repulsion between like-charged particles that is proportional to $1/r^3$, so overall there is no attraction at all. The capillary deformation of the interface contributes a mere 1.8×10^{-5} kT to the interaction potential, so it is insignificant thermodynamically. Thus, the mechanism proposed by Nikolaides *et al.*¹ does not account for the observations, and the origin of the observed attraction remains enigmatic.

Mischa Megens*, Joanna Aizenberg†

*Philips Research, 5656 AA Eindhoven,

The Netherlands

e-mail: mischa.megens@philips.com

†Bell Laboratories/Lucent Technologies,

Murray Hill, New Jersey 07974-2008, USA

1. Nikolaides, M. G. *et al.* *Nature* **420**, 299–301 (2002).

2. Stamou, D., Duschl, C. & Johannsmann, D. *Phys. Rev. E* **54**, 5263–5272 (2000).

Nikolaides et al. reply — Our experiment¹ provides a clear measure of the interactions between charged particles at fluid–fluid interfaces and demonstrates that there can be a long-range attractive interaction between such particles.

However, Megens and Aizenberg raise important points about our interpretation of these results. Our calculations account for the electrostatic stresses acting on the fluid–fluid interface, but neglect the force that the electric field exerts on the particle itself. A detailed evaluation of this force, obtained by calculating the change in electrostatic energy as the particle is pushed into the fluid, shows that the interfacial force pulling the particle out of the fluid is exactly cancelled by the electrical force pushing the particle into the fluid, in agreement with the suggestion of Megens and Aizenberg.

This creates a puzzle: the data show unambiguously that there is a long-range attraction, but what is its origin? The particles have a long-range repulsive interaction, owing to their charges; this is dipolar in character, with the electrostatic energy decaying as $1/r^3$. The attractive interaction must balance this electrostatic repulsion, so, if it decays as a power law, the attractive interaction energy must decay more slowly than $1/r^3$ to create the stable energy minimum observed experimentally. This eliminates possibilities such as asymmetries in the contact line or fluctuation-induced forces, all of which have a power-law decay that is more rapid than $1/r^3$. The most likely interaction that has sufficient range therefore remains capillary distortion of the interface.

Capillary distortion can occur only if there is an imbalance between the force pulling the particle into the water and the force pushing the interface outwards towards the oil. However, it is sufficient for these forces to be imbalanced only up to distances comparable to the interparticle separation: this would still lead to distortion of the interface between the particles, causing the attractive force. The resolution of the puzzle could lie in charges on the particles on the oil side, rather than on the water side². As the density of free charges in oil is much lower than that in water, the screening length is correspondingly larger; this significantly extends the range of the force imbalance and can account for our experimental observations. The origin of the capillary distortion therefore remains electrostatic in nature.

We have since confirmed that the particles do have a measurable charge, even when immersed in oil. Measurements of the electrophoretic mobility and pair correlation functions of these particles indicate that their charge can be as high as $200e$ in oil. In the absence of additional solubilized charges, this would result in a screening length of the order of $20 \mu\text{m}$ for the particle volume fraction of about 10^{-4} used in our experiments. This is larger than the particle separation, allowing the force imbalance between the particle and the interface to persist far enough for significant interfacial distortion to exist at scales comparable to the interparticle separation. We therefore believe that electric-field-induced capillary distortion remains the likely culprit for the attractive interactions between like-charged interfacial particles.

M. G. Nikolaides, A. R. Bausch, M. F. Hsu,

A. D. Dinsmore, M. P. Brenner, C. Gay,

D. A. Weitz*

*Department of Physics, and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
e-mail: weitz@deas.harvard.edu

1. Nikolaides, M. G. *et al.* *Nature* **420**, 299–301 (2002).

2. Aveyard, R. *et al.* *Phys. Rev. Lett.* **88**, 246102-1-4 (2002).

Electric field effect in correlated oxide systems

C. H. Ahn¹, J.-M. Triscone² & J. Mannhart³

¹Department of Applied Physics, Yale University, PO Box 208284, New Haven, Connecticut 06520-8284, USA

²Condensed Matter Physics Department, University of Geneva, 24 quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland

³Center for Electronic Correlations and Magnetism, Institute of Physics, University of Augsburg, D-86135 Augsburg, Germany

Semiconducting field-effect transistors are the workhorses of the modern electronics era. Recently, application of the field-effect approach to compounds other than semiconductors has created opportunities to electrostatically modulate types of correlated electron behaviour—including high-temperature superconductivity and colossal magnetoresistance—and potentially tune the phase transitions in such systems. Here we provide an overview of the achievements in this field and discuss the opportunities brought by the field-effect approach.

Our daily life is permeated by semiconducting field-effect transistors (FETs). About 10^{18} of these microscopic electronic switches are produced every year to run the tools that are indispensable to us, including our cars, computers, cellular phones and kitchen appliances. These devices work by modulating the electrical charge carrier density, and hence electrical resistance, of a thin semiconducting channel through the application of an electric field. This remarkably simple and very successful principle provides new opportunities for basic science and innovative device applications when applied to novel correlated electron systems whose properties depend strongly on the carrier concentration. Excited by the opportunities of the field effect in such materials—including organic conductors^{1,2}, high-temperature superconductors^{3,4} and colossal magnetoresistance compounds⁵—dozens of groups have worked on this approach, yielding a panoply of striking results⁶. The burgeoning field of plastic electronics is a notable example that has reached technological fruition².

These opportunities probably served as a motivation for the research directions of the group at Bell Laboratories⁷. Several of this group's articles in *Nature* and *Science*, for example, addressed electric-field-induced superconductivity in new materials. The public debate about the scientific misconduct that underlies a

large part of this group's claims⁷, however, may unfortunately and wrongly give the impression that the electric field effect in such materials does not exist. Here we provide an overview of the electrostatic field effect in complex oxides, one of the newest classes of materials to which the field effect has been applied. We consider the role of electronic correlations, discuss examples in superconductivity and magnetism, and outline future directions, commenting on the technical challenges of this approach.

Strongly correlated systems

In materials ranging from ionic insulators and elemental semiconductors to simple metals, the band structure model and associated Fermi liquid description provide a deep understanding of their fundamental physical properties. In this framework, insulators and intrinsic semiconductors have only filled and empty bands, whereas metals have at least one partially filled band. These concepts reach their limits, however, as the interactions between electrons become strong or the dimensionality of the system is reduced. When the electrostatic Coulomb interactions become dominant, localization occurs for commensurate filling (for example, 1/4 or 1/2 filling), resulting in a special type of insulator, frequently called the Mott insulator⁸. One-dimensional (1D) GaAs quantum wires, organic

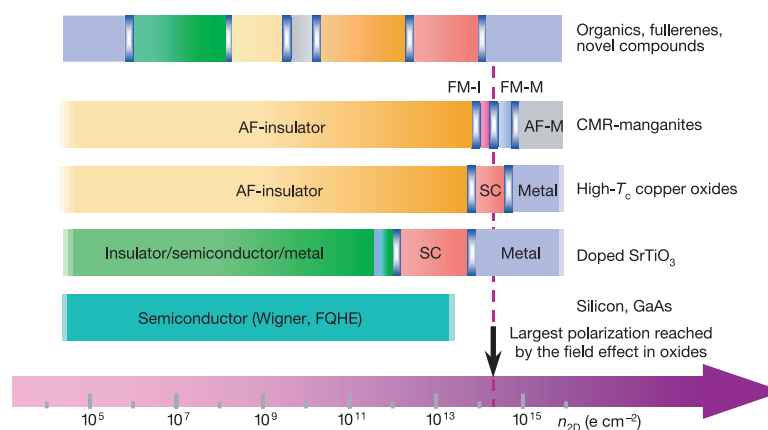


Figure 1 Illustration of the zero-temperature behaviour of various correlated materials as a function of sheet charge density (n_{2D}). Silicon is shown as a reference. The examples for high- T_c superconductors and for colossal magnetoresistive (CMR) manganites reflect $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ and $(\text{La}, \text{Sr})\text{MnO}_3$, respectively. The top bar has been

drawn to illustrate schematically the richness of materials available for field-effect tuning and the spectrum of their phases. AF, antiferromagnetic; FM, ferromagnetic; I, insulator; M, metal; SC, superconductor; FQHE, fractional quantum Hall effect; Wigner, Wigner crystal.

Bechgaard salts, and undoped two-dimensional (2D) high-temperature superconductors are notable examples of such insulators^{1,9,10}. Disorder in these correlated electron systems also has an important role. At low densities, depending on the balance between disorder and the strength of the electronic correlations, one expects to have Anderson localization (in 1D or 2D) or a crystallization of the electrons into a Wigner crystal; both schemes lead to an insulating state^{11,12}. Recent work on 2D, high-mobility Si field-effect structures proposes that extremely strong electron interactions may overcome the disorder that leads to localization, resulting in a metallic ground state¹³.

Interactions combined with low dimensionalities thus lead to novel phases, which are strongly influenced by the carrier concentration¹⁴. Figure 1 illustrates several of the properties of the materials discussed above as a function of their sheet charge density; silicon is shown as a reference. Because the carrier concentration is a crucial parameter that governs the properties of these systems, the electrostatic field-effect approach is an ideal tool to investigate the physics of correlated electron systems, allowing controlled and reversible changes of the carrier concentration without altering the disorder. The fractional quantum Hall effect, exhibited in GaAs/AlGaAs heterostructures, is a striking example of correlations in a clean, 2D system, in which the field effect has been used effectively to tune the carrier density¹⁵. Other methods to change the carrier density without affecting the disorder include the application of pressure and photon irradiation¹⁶. One challenge that is apparent from Fig. 1, however, is that much of the novel behaviour beyond the realm of the fractional quantum Hall effect and Wigner crystallization occurs at relatively high carrier densities, requiring the use of extremely large polarization fields for the field effect, as described below.

Experimental parameters

Field-effect experiments are conceptually elegant, but pose significant challenges in practice because the underlying physics requires structures that are difficult to achieve. Applying an external electric field to a material attracts or repels charge carriers, creating a thin charge accumulation or depletion layer at the surface that modifies the electrical conductivity between a source (S) and a drain (D) contact (Fig. 2). The field is applied across a gate insulator using a gate electrode (G). The characteristic width of the accumulation or depletion layer is given by the electrostatic screening length λ_{el} , which in the semiclassical, metallic limit is the Thomas–Fermi length. In standard metallic systems, λ_{el} is extremely short, a fraction of an atomic diameter, and thus negligible field effects are found. In low-carrier-density systems, larger screening lengths and field effects are expected, which are reasons why standard transistors are made with semiconductors rather than metals.

Many of the interesting physical properties described in Fig. 1 occur at carrier densities in the range of 10^{19} – 10^{22} carriers cm^{-2} , intermediate between those of conventional metals and semiconductors such as Si or GaAs. To achieve substantial carrier modulation at these densities, two approaches are followed. The first is to use ultrathin drain–source (DS) channels, which are only a few nanometres thick, so that the absolute number of carriers is small, $\sim 10^{14} \text{ cm}^{-2}$, allowing large relative changes in their total number. Because the carrier modulation occurs in an interfacial layer, it is also advantageous to use ultrathin channels to avoid shunting effects. The second means of achieving large modulation of the carrier density is to apply extremely large electric fields across insulators that are tailored for large breakdown field strengths (E_{b}) and dielectric constants (ϵ_{r}). The polarization (σ), or areal charge density (charge cm^{-2}), that can be induced at a given gate voltage V_{G} applied across an insulator of thickness t is given by $\sigma = \int_0^{V_{\text{G}}} \epsilon_0 \epsilon_{\text{r}}(V)/t \, dV$, where $V_{\text{G}} < E_{\text{b}}t$. To change the carrier density by 50% in a 1-nm-thick oxide superconductor with 4×10^{21} holes cm^{-3} , for example, requires sheet charge densities or polarizations of $30 \mu\text{C cm}^{-2}$, an order of magnitude larger than what can be achieved in standard FETs.

In addition to this stringent requirement on the sustainable polarization, leakage currents through the insulator must be much smaller than the DS current, and the density of localized interface states must be small compared with the modulated carrier density. In conventional Si FETs, the universal gate dielectric is SiO_2 , an excellent insulator with low leakage that is chemically stable and virtually free of interface states. These outstanding properties of SiO_2 are a primary reason for the success of Si-based electronics. For oxide FETs, however, SiO_2 is unsuitable because its breakdown field $E_{\text{b}} = 10 \text{ MV cm}^{-1}$ and dielectric constant $\epsilon_{\text{r}} = 3.9$ lead to polarizations of $3 \mu\text{C cm}^{-2}$ (2×10^{13} charges cm^{-2}) at best. In the case of plastic electronics, this relatively small polarization does not preclude the use of SiO_2 , although issues related to traps, surface states, degradation of the channel, and leakage had to be mastered. An additional difficulty for organic FETs is that the design and fabrication of the contacts to the DS channels are non-trivial.

Because of these technical challenges, chemical doping, which can achieve metallic carrier densities, is by far the most commonly used method to manipulate the carrier densities of complex oxides. Largely owing to advances in physical vapour deposition techniques, it is only in recent years that doping of complex oxides by the electric field effect has become possible. It is now routine to fabricate epitaxial high- ϵ_{r} complex oxide dielectrics, such as SrTiO_3 , and ferroelectric oxides, such as $\text{Pb}(\text{Zr,Ti})\text{O}_3$ (PZT), that can achieve polarizations in the range of 10 – $40 \mu\text{C cm}^{-2}$. *In situ* multi-layer growth of the oxide dielectric and DS channels has led to improved structural and electronic properties of the DS channel with acceptable interface trap densities^{6,17,18}.

Interestingly, these advances in complex oxide gate dielectrics have led the semiconductor industry to consider replacing SiO_2 with high- ϵ_{r} complex oxide materials. As the thickness of the SiO_2 layer approaches one nanometre, below which leakage from tunnel currents becomes unacceptably large, alternative gate dielectrics will become a necessity¹⁹. The possibility of using thicker gate dielectrics with larger values of ϵ_{r} is driving the exploration of complex oxide gate dielectrics, as well as the development of the methodologies to deposit these materials directly on Si. As with complex oxide FETs, challenges related to the density of interface states and the chemical and thermodynamic compatibility of the gate insulator with the DS-channel material are being addressed^{20–22}.

Despite the technical challenges of fabricating complex oxide FETs, the notion of switching exotic behaviour (such as superconductivity) on and off just by turning the knob on a power supply is so tempting that dozens of groups have addressed these obstacles, with experiments starting as early as 1960 (for example, refs 23–34; for an extended list, see <http://www.physik.uni-augsburg.de/exp6/field-effect/>). It was in 1991, working with high superconducting transition temperature (high T_{c}) copper oxide superconductors, that Mannhart and colleagues at the IBM Research Laboratory in Rüschlikon observed electric-field-induced shifts of T_{c} of several K in epitaxial heterostructures (Fig. 3a)²⁵. Subsequent work using the ferroelectric field effect has also allowed reversible modulation of superconductivity to be induced in these materials¹⁸ (Fig. 3b, c).

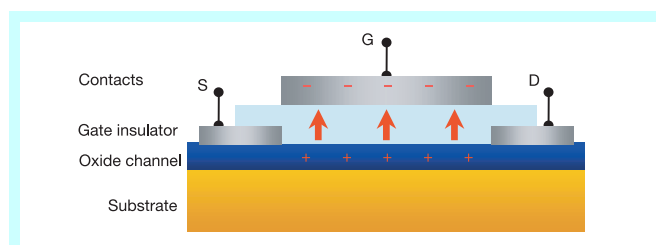


Figure 2 Cross-section of a typical sample geometry used for field-effect studies. S, source; G, gate; D, drain.

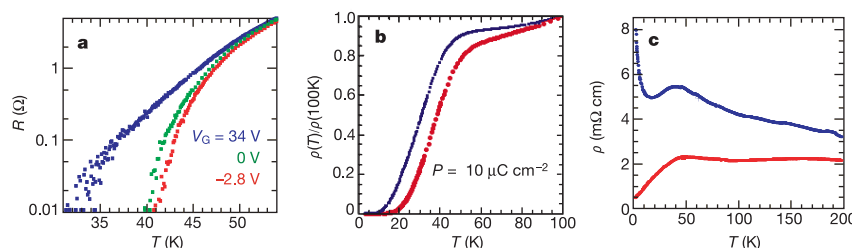


Figure 3 Field effects in superconducting films. In each case, the blue curve corresponds to depletion of the carrier density, and the red curve corresponds to enhancement of the carrier density in the DS (drain–source) channel. **a**, Change of the DS resistance of an ~ 8 -nm-thick $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ channel with a ~ 300 -nm-thick $\text{Ba}_{0.15}\text{Sr}_{0.85}\text{TiO}_3$ gate insulator. The scatter in the data results from the noise of the measurement system (from ref. 6). **b**, Resistance change of a ~ 2 -nm-thick

$\text{GdBa}_2\text{Cu}_3\text{O}_{7-\delta}$ film induced by a 300-nm-thick PZT layer acting as ferroelectric gate. The two curves have been normalized in the normal state. **c**, Resistance change of a ~ 2 -nm-thick $\text{GdBa}_2\text{Cu}_3\text{O}_{7-\delta}$ film whose doping level has been chosen to be close to the superconductor–insulator transition, induced by a 300-nm-thick PZT layer acting as ferroelectric gate. These data have been measured at 1 T (from ref. 18).

Correlated oxide superconductivity

We now turn to the role of charge in the physics of the high- T_c copper oxide superconductors. In transition metal oxides with a partially filled d -electron band, the electronic correlations lead to the opening of a direct gap in the $3d$ band (Mott–Hubbard insulators) or an indirect gap between the oxygen $2p$ band and the upper Hubbard band (charge transfer insulators); both cases result in an insulating state at half filling. The high- T_c superconductors are doped charge transfer insulators, and the role of carriers is depicted in the temperature–doping (T – x) phase diagram of Fig. 4. The parameter x stands for the doping level, which is usually changed by non-isovalent chemical substitution of the insulating $x = 0$ parent compound, and is roughly proportional to the number of ‘free’ electrical carriers in the system. At low doping levels, the system is an antiferromagnetic insulator owing to strong electronic correlations. As the doping level is increased, superconductivity appears, with the transition temperature T_c following a roughly parabolic dependence on x .

At low doping levels, near the superconductor–insulator phase boundary, field-effect tuning of the carrier density should induce a reversible insulator–superconductor quantum phase transition, an active area of research in condensed matter physics^{35,36}. Such superconducting switching behaviour is appealing from both a fundamental perspective and for potential applications. Figure 3c shows ferroelectric field-effect experiments on underdoped, ultrathin (20 Å) $\text{GdBa}_2\text{Cu}_3\text{O}_7$ films, in which non-volatile, reversible switching between superconducting and insulating behaviour could be achieved by modulating the polarization state of the ferroelectric gate¹⁸. This switching occurs at a very large value of the sheet resistance (about 40 k Ω per square), in agreement with values obtained by chemical doping and disorder-induced superconductor–insulator transitions in high-temperature superconductors, but significantly larger than the values observed in low-temperature superconductors. On the insulating side, analysis of the transport properties reveals a resistivity described by $\rho \propto \ln(1/T)$. This peculiar temperature dependence was also observed in very-high-magnetic-field (60 T) experiments that allowed superconductivity to be suppressed in underdoped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ and other compounds³⁷. These results suggest that the physics of the insulating state reached by high electric or magnetic fields is different from that obtained by chemical doping, which may be related to the fact that even as the superfluid density is modulated by the applied electric or magnetic field, the microstructure and disorder remain fixed.

In addition, in the underdoped region of the T – x phase diagram, the copper oxide superconductors have a relatively low carrier density, resulting in a low superfluid density and weak phase stiffness. It has thus been proposed that the T_c in this region is controlled by phase fluctuations^{38,39}. As a result, T_c only reflects the

establishment of long-range phase coherence and should be proportional to the superfluid density n_s . This relationship has been observed experimentally in muon spin resonance experiments on chemically doped bulk samples⁴⁰. Recent ferroelectric field-effect experiments on underdoped thin films of $\text{NdBa}_2\text{Cu}_3\text{O}_{7-\delta}$ have also correlated quantitatively changes in T_c with changes in the carrier concentration produced by the field effect, in agreement with the phase fluctuation model⁴¹. In contrast to chemical doping, the field-effect experiments only modify the charge, revealing directly the relationship between the carrier density and T_c .

Electrostatic modulation of the carrier density has also been used to study the normal state of the copper oxide superconductors above T_c . Besides the opening of a gap in the electronic spectrum (called the pseudogap) at a temperature T^* (Fig. 4), it was recognized early on that the normal-state transport properties are anomalous, in particular the temperature dependence of the resistivity and Hall constant R_H . In ferroelectric/superconducting heterostructures, analyses of the normal-state temperature dependence of the resistivity and Hall constant as a function of the ferroelectric polarization show that both these quantities can be rescaled over the entire measured temperature range⁴². These results point to a temperature-indepen-

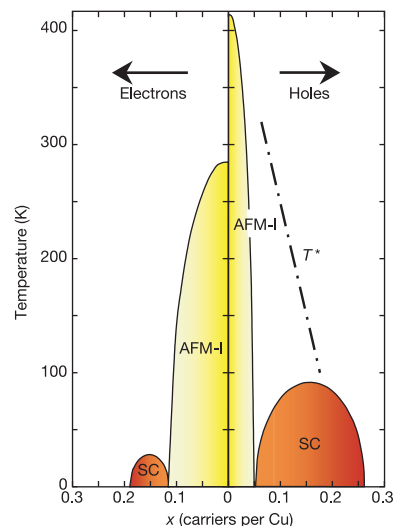


Figure 4 Phase diagram of the high- T_c superconductors. The electron-doped material represents $\text{Nd}_{2-x}\text{Ce}_x\text{CuO}_{4-y}$, and the hole-doped system represents $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. The insulating, antiferromagnetic phases (AFM-I) are drawn in yellow, and the superconducting phases (SC) are in orange.

dent carrier concentration, with the observed temperature dependence of the Hall constant possibly being related to the temperature dependences of the scattering times and effective masses.

New approaches

Although most field-effect work on complex oxides has focused on superconductors, the approach has been extended recently to other materials. Manganites exhibiting colossal magnetoresistance (CMR) reveal large changes of their magnetic and transport properties induced by electric fields^{43–45}. These experiments show that charge modulation at fixed disorder influences the CMR effect, whereas existing theoretical models have highlighted the role of disorder. One can also consider the possibility of achieving electric-field-tuned metal–insulator phase transitions, an idea that is being explored for applications based on oxide channel FETs operating at room temperature^{46,47}. In addition, the field effect in perovskite-based organic–inorganic hybrid materials is being examined to further improve the performance of low-cost thin-film transistors². In these approaches, materials preparation remains a key challenge. As with the development of semiconductor heterostructures, which took many years but has provided spectacular dividends, it will take a sustained effort—including the development of new fabrication techniques and advanced nanoscale characterization tools—to spur further advances in the observation of novel field effects in correlated oxide systems.

What are the future prospects for the field? The diversity of behaviour exhibited by new materials with correlated electrons is so rich that the field effect offers many possibilities. In particular, the polarization fields that are now routinely achievable using ferroelectrics and high- ϵ_r complex oxides provide access to largely uncharted territory, where the interplay between charge at high carrier densities in clean systems and strong Coulomb repulsion becomes important. Nanoscale control of electronic properties using field effects is also an emerging area. Recently, scanning probe microscopy has been used to induce localized field effects in magnetic and superconducting oxides^{42,48}. The possibility of controlling the electronic properties of materials with nanometre resolution will allow the examination of fundamental issues underlying the behaviour of correlated oxides, such as nanoscale electronic phase separation⁴⁹. It would also facilitate the optimization of physical properties through reversible nanostructuring. In future superconducting devices, nanoscale pinning sites could be introduced, and Josephson junctions and arrays of junctions could be written into superconductors. Combined with engineering breakthroughs taking place in nanotechnology, such as the fabrication of high-density arrays of scanning probes⁵⁰, novel device and memory structures could be fabricated. The field effect, which has been so spectacularly successful in its application to semiconductors, is only in its infancy in exotic correlated systems; the combination of advanced twenty-first-century techniques with this new materials approach offers broad perspectives for basic science and technology. □

doi:10.1038/nature01878.

- Jérome, D. & Caron, L. G. (eds) *Low-Dimensional Conductors and Superconductors* (NATO Advanced Study Institutes, Ser. B, Vol. 155, Plenum, New York, 1987).
- Shaw, J. M. & Seidler, P. F. (eds) *Organic electronics*. *IBM J. Res. Dev.* **45**(1), 3–128 (2001).
- Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64**, 189–193 (1986).
- Wu, M. K. *et al.* Superconductivity at 93 K in a new mixed-phase Y-Ba-Cu-O compound system at ambient pressure. *Phys. Rev. Lett.* **58**, 908–912 (1987).
- Tokura, Y. (ed.) *Colossal Magnetoresistive Oxides* (Advances in Condensed Matter Science, Vol. 2, Gordon and Breach, London, 2000).
- Mannhart, J. High- T_c transistors. *Supercond. Sci. Technol.* **9**, 49–67 (1996).
- Results of Inquiry Into the Validity of Certain Physics Research Papers from Bell Labs*. Available at (http://www.lucnet.com/news_events/researchreview.html) (2002).
- Mott, N. *Metal-Insulator Transitions* (Taylor and Francis, London, 1990).
- de Picciotto, R. *et al.* 2D–1D coupling in cleaved edge overgrowth. *Phys. Rev. Lett.* **85**, 1730–1733 (2000).
- Orenstein, J. & Millis, A. J. Advances in the physics of high-temperature superconductivity. *Science* **288**, 468–474 (2000).
- Abrahams, E., Kravchenko, S. V. & Sarachik, M. P. Colloquium: Metallic behavior and related phenomena in two dimensions. *Rev. Mod. Phys.* **73**, 251–266 (2001).

- Wigner, E. On the interaction of electrons in metals. *Phys. Rev.* **46**, 1002–1011 (1934).
- Kravchenko, S. V. *et al.* Scaling of an anomalous metal–insulator transition in a two dimensional system in silicon at $B = 0$. *Phys. Rev. B* **51**, 7038–7045 (1995).
- Imada, M., Fujimori, A. & Tokura, Y. Metal–insulator transitions. *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
- Stormer, H. L., Tsui, D. C. & Gossard, A. C. The fractional quantum Hall Effect. *Rev. Mod. Phys.* **71**, S298–S305 (1999).
- Chu, C. W. Materials and physics of high temperature superconductors: A summary, two recent experiments and a comment. *Physica Scripta T* **102**, 40–50 (2002).
- Christen, H.-M., Mannhart, J., Williams, E. J. & Gerber, Ch. Dielectric properties of sputtered SrTiO₃ films. *Phys. Rev. B* **49**, 12095–12104 (1994).
- Ahn, C. H. *et al.* Electrostatic modulation of superconductivity in ultra-thin GdBa₂Cu₃O₇ films. *Science* **284**, 1152–1155 (1999).
- The International Technology Roadmap for Semiconductors 2002. Available at (<http://public.itrs.net/>) (2002).
- Hubbard, K. J. & Schlom, D. G. Thermodynamic stability of binary oxides in contact with silicon. *J. Mater. Res.* **11**, 2757–2776 (1996).
- McKee, R. A., Walker, F. J. & Chisholm, M. F. Crystalline oxides on silicon: The first five monolayers. *Phys. Rev. Lett.* **81**, 3014–3017 (1998).
- Lin, A. *et al.* Epitaxial growth of Pb(Zr_{0.2}Ti_{0.8})O₃ on Si and its nanoscale piezoelectric properties. *Appl. Phys. Lett.* **78**, 2034–2036 (2001).
- Glover, R. E. & Sherill, M. D. Changes in superconducting critical temperature produced by electrostatic charging. *Phys. Rev. Lett.* **5**, 248–250 (1960).
- Hebard, A. F. & Fiory, A. T. in *Novel Superconductivity* (eds Wolf, S. A. & Kresin, V. Z.) **9** (Plenum, New York, 1987).
- Mannhart, J., Bednorz, J. G., Müller, K. A. & Schlom, D. G. Electric field effect on superconducting YBa₂Cu₃O_{7- δ} films. *Z. Phys. B* **83**, 307–311 (1991).
- Levy, A. *et al.* Field and Hall effects in semiconducting YBa₂Cu₃O_{6+ δ} . *Phys. Rev. B* **46**, 520–523 (1992).
- Xi, X. X. *et al.* Effects of field-induced hole-density modulation on normal-state and superconducting transport in YBa₂Cu₃O_{7- δ} . *Phys. Rev. Lett.* **68**, 1240–1243 (1992).
- Ivanov, Z. G., Stepanov, E. A., Tzalenchuk, A., Ya, Shekhter, R. I. & Claeson, T. Field effect transistor based on a bi-crystal grain boundary Josephson junction. *IEEE Trans. Appl. Supercond.* **3**, 2925–2927 (1993).
- Joesse, K. *et al.* Electric-field effect devices made of YBa₂Cu₃O_{7- δ} /SrTiO₃ epitaxial multilayers. *Physica C* **224**, 179–184 (1994).
- Matijasevic, V. C. *et al.* Electric field induced superconductivity in an overdoped copper oxide superconductor. *Physica C* **235–240**, 2097–2098 (1994).
- Mannhart, J., Ströbel, J., Bednorz, J. G. & Gerber, Ch. Large electric field effects in YBa₂Cu₃O_{7- δ} films containing weak links. *Appl. Phys. Lett.* **62**, 630–632 (1993).
- Mayer, B., Mannhart, J. & Hilgenkamp, H. Electric-field controllable Josephson junctions of high quality in high- T_c superconductors. *Appl. Phys. Lett.* **68**, 3031–3033 (1996).
- Watanabe, Y. Epitaxial all-perovskite field effect transistor with a memory retention. *Appl. Phys. Lett.* **66**, 1770–1772 (1995).
- Aidam, R., Fuchs, D. & Schneider, R. Ferroelectric field effect in YBa₂Cu₃O_{7- δ} thin films. *Physica C* **328**, 21–30 (1999).
- Goldman, A. M. & Markovic, N. Superconductor–insulator transitions in the two-dimensional limit. *Phys. Today* **51**, 39–44 (1998).
- Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, Cambridge, 1999).
- Boebinger, G. S. *et al.* Insulator-to-metal crossover in the normal state of La_{2-x}Sr_xCuO₄ near optimum. *Phys. Rev. Lett.* **77**, 5417–5420 (1996).
- Emery, V. J. & Kivelson, S. A. Superconductivity in bad metals. *Phys. Rev. Lett.* **74**, 3253–3256 (1995).
- Schneider, T. & Singer, J. M. Crossover and quantum critical phenomena in high temperature superconductors. *Europhys. Lett.* **40**, 79–84 (1997).
- Uemura, Y. J. *et al.* Magnetic field penetration depth in Tl₂Ba₂CuO_{6+ δ} in the overdoped regime. *Nature* **364**, 605–607 (1993).
- Gariglio, S., Ahn, C. H., Triscone, J.-M. Electrostatic modulation of the superconducting transition in thin NdBa₂Cu₃O_{7- δ} films: The role of classical fluctuations. *Physica C* **372–376**, 583–586 (2002).
- Gariglio, S., Ahn, C. H., Matthey, D. & Triscone, J.-M. Electrostatic tuning of the hole density in NdBa₂Cu₃O_{7- δ} films and its effect on the Hall response. *Phys. Rev. Lett.* **88**, 67002 (2002).
- Mathews, S., Ramesh, R., Vankatesan, T. & Benedetto, J. Ferroelectric field effect transistor based on epitaxial perovskite heterostructures. *Science* **276**, 238–240 (1997).
- Tanaka, H., Zhang, J. & Kawai, T. Giant electric field modulation of double exchange ferromagnetism at room temperature in the perovskite manganite/titanate p-n junction. *Phys. Rev. Lett.* **88**, 027204 (2002).
- Hong, X., Posadas, A. & Ahn, C. H. Structural and electronic properties of PZT/ultrathin La_{0.7}Sr_{0.3}MnO₃ heterostructures for ferroelectric field effect studies. *Solid State Electron* (in the press).
- Newns, D. M. *et al.* Mott transition field effect transistor. *Appl. Phys. Lett.* **73**, 780–782 (1998).
- Ueno, K. *et al.* Field-effect transistor on SrTiO₃ with sputtered Al₂O₃ gate insulator. *Proc. M²S 2003 Conf.* (in the press); preprint at (<http://xxx.lanl.gov/cond-mat/0306436>) (2003).
- Ahn, C. H. *et al.* Local nonvolatile electronic writing of epitaxial Pb(Zr,Ti)O₃/SrRuO₃ heterostructures. *Science* **276**, 1100–1103 (1997).
- Mathur, N. & Littlewood, P. Mesoscopic texture in manganites. *Phys. Today* **56**, 25–30 (2003).
- Levi, B. G. Do atomic force microscope arrays have the write stuff? *Phys. Today* **55**, 14–17 (2002).

Acknowledgements We acknowledge discussions with, and support from, T. Giamarchi, A. Lin, A. Sawa, D. G. Schlom, C. W. Schneider and C. C. Tsuei. C.A. was supported by AFOSR, DOE, NSF and the Packard Foundation; J.M. was supported by BMBF, ESF (Thiox) and SFB; J.M.T. was supported by the Swiss National Science Foundation through the National Centre of Competence in Research, ‘Materials with Novel Electronic Properties’ and division II, NEDO, and ESF (Thiox).

Correspondence and requests for materials should be addressed to C.H.A. (charles.ahn@yale.edu), J.M.T. (jean-marc.triscone@physics.unige.ch) or J.M. (jochen.mannhart@physik.uni-augsburg.de).

A strong astrophysical constraint on the violation of special relativity by quantum gravity

T. Jacobson, S. Liberati & D. Mattingly

Department of Physics, University of Maryland, College Park, Maryland 20742-4111, USA

Special relativity asserts that physical phenomena appear the same to all unaccelerated observers. This is called Lorentz symmetry and relates long wavelengths to short ones: if the symmetry is exact it implies that space-time must look the same at all length scales. Several approaches to quantum gravity, however, suggest that there may be a microscopic structure of space-time that leads to a violation of Lorentz symmetry. This might arise because of the discreteness¹ or non-commutativity² of space-time, or through the action of extra dimensions³. Here we determine a very strong constraint on a type of Lorentz violation that produces a maximum electron speed less than the speed of light. We use the observation of 100-MeV synchrotron radiation from the Crab nebula to improve the previous limit by a factor of 40 million, ruling out this type of Lorentz violation, and thereby providing an important constraint on theories of quantum gravity.

To characterize Lorentz violation we adopt the simple framework of deformed dispersion relations. A dispersion relation is the relation between energy E and momentum p for a particle. In relativity, these quantities change under a Lorentz transformation, but the particle dispersion relation $E^2 = m^2c^2 + p^2c^4$ is invariant (where m is the particle mass and c is the speed of light). Lorentz transformations include both 'boosts', which are transformations to a relatively moving frame, and rotations.

We consider dispersion relations that break boost invariance but preserve rotation invariance. (A dispersion relation that is not boost invariant can hold in only one frame. We assume that this frame coincides with that of the cosmic microwave background.) In particular for photons and electrons we consider the dispersion relations:

$$\omega^2(k) = k^2 + \xi \frac{k^3}{M} \quad (1)$$

$$E^2(p) = m^2 + p^2 + \eta \frac{p^3}{M} \quad (2)$$

where ω and k are the photon frequency and wavenumber, and E and p are the electron energy and momentum. We use units with Planck's constant $\hbar$ and the low energy speed of light c both set to unity (thus the photon energy $\hbar\omega$ is also denoted by ω). The energy scale $M = 10^{19}$ GeV, which is close to the Planck energy, $M_{\text{Planck}} = 1.22 \times 10^{19}$ GeV, is factored out explicitly. We regard these relations as phenomenological, not fundamental. In particular, one would expect a whole series of higher order terms in the momentum, but those would be negligible at the energies we consider. Extra lower order terms are already strongly enough constrained by observation so as to be irrelevant for the new constraint considered here.

Dispersion relations with higher order momentum corrections analogous to equation (1) are familiar from common physical situations like vibrational waves in a crystal or light waves in a refractive medium, where the long wavelength (low-momentum) modes travel at a common speed, while the modes whose wavelength is short enough to be sensitive to the microscopic structure of the medium behave differentially. Several approaches to quantum

gravity propose a Lorentz violating microscopic structure of space-time. There is no unique prediction regarding the values of the dimensionless coefficients ξ and η , but if not suppressed by some other symmetry they would presumably be of order unity if they originate from quantum gravity effects. This is because physics rarely produces dimensionless numbers that differ by many orders of magnitude from unity from theories with only one scale, and the unique energy scale M of quantum gravity has already been factored out in equations (1) and (2).

It is of course possible that other symmetries along with boost symmetry are violated. However, a bound on pure boost violation is also a bound on such multiple symmetry violations, barring an unlikely cancellation of different effects. Hence, to keep the analysis simple we assume all of standard physics except boost invariance. In particular, we preserve rotation symmetry and electromagnetic gauge invariance, additivity of energy and momentum for multi-particle systems, and energy-momentum conservation in particle interactions.

A consistent dynamical framework that yields modified dispersion and preserves our other assumptions is effective field theory. Effective field theory is very general, and can incorporate Lorentz violation arising from a wide range of underlying quantum gravity scenarios, including for example string theory and space-time discreteness. It has been shown⁴ in effective field theory that left and right polarizations of photons have opposite values of ξ in equation (1), while left and right electron chiralities can have independent η values. Polarization dependence of ξ is unimportant for the new constraint described here, although we shall use it at the end when combining our new constraint with previous work. We assume here that η is not chirality dependent.

Perhaps surprisingly, Planck energies are not needed to get constraints on ξ and η of order unity or even less. For example, the group velocities $d\omega/dk$ and dE/dp for photons and electrons satisfying equations (1) and (2) depend on these parameters and on the particle energy. Over long propagation distances this could produce observable arrival time differences for simultaneously emitted photons of different energies^{5,6}. Arrival time constraints have been obtained using the observed radiation from γ -ray bursts⁷, active galactic nuclei⁸, and pulsars⁹. The best that can be done with the current data is $|\xi| \lesssim O(100)$ (refs 8, 9), but more stringent limits can be expected in the future¹⁰. Constraints of order 10^{-4} have been obtained on the difference in ξ for different polarizations (birefringence) using spectropolarimetric observations of distance galaxies¹¹. Using the effective field theory result of ref. 4, this yields the powerful new constraint $|\xi| \lesssim O(10^{-4})$.

Complementary constraints were recently achieved by considering the threshold reactions of photon decay, vacuum Čerenkov radiation, and photon absorption involving high energy photons and electrons¹²⁻¹⁹. Energies around 10 TeV are needed in order to put constraints or order unity on the deformation parameters for electrons and photons using threshold effects. Energies higher than this are achieved in astrophysical sources, and hence threshold reactions place at least order unity bounds on η , ξ .

However, if we are to rule out quantum gravity effects characterized by equations (1) and (2), it is necessary to find observations that limit η and ξ to be much smaller than unity. None of the previous work has accomplished this for both electrons and photons. Here we show that the presence of synchrotron radiation from the Crab nebula, emitted by electrons cycling in a magnetic field, provides the new constraint $\eta > -7 \times 10^{-8}$ on the electron parameter η within the effective field theory framework. This improves previous constraints by a very large amount. In particular, it implies that the parameter E_{QG} of the popular phenomenological model of ref. 7 (which corresponds to $M/|\eta|$, with $\xi = \eta$ in our notation) is bounded by $E_{\text{QG}} > 10^7 M_{\text{Planck}}$. (To be compatible

with effective field theory that model must be modified to accommodate the polarization-dependence of the sign of ξ .) This is seven orders of magnitude stronger than the constraint from photon absorption $E_{\text{QG}} > 0.3 M_{\text{Planck}}$ (refs 15, 16, 20) (see also refs 18 and 19 for a debate on the validity of this constraint) and three orders of magnitude stronger than the new birefringence constraint of ref. 4.

Synchrotron radiation in the Crab nebula is produced by electrons cycling in a magnetic field of around 0.6 mG. To produce the observed radiation of energy 100 MeV in this field requires in standard electrodynamics a gamma factor $\gamma = (1 - v^2)^{-1/2}$ of 3×10^9 , corresponding to an electron energy of 1,500 TeV. To achieve this energy, the electron velocity must differ from the speed of light by less than one part in 10^{19} . Lorentz violation with negative η puts an upper bound on the electron speed, and hence an upper bound on the possible synchrotron radiation frequency. (The existence of a cut-off in the synchrotron radiation at high charged particle energies in the presence of Lorentz violation was previously noted in ref. 21.) But to apply this reasoning, we must first re-analyse the synchrotron emission process allowing for the Lorentz violating effects that we wish to bound.

In standard electrodynamics, accelerated electrons in a magnetic field emit synchrotron radiation with a spectrum that sharply cuts off at a frequency ω_c given by the formula²²

$$\omega_c^{\text{LI}} = \frac{3eB\gamma^2}{2m} \quad (3)$$

where B is the component of the magnetic field orthogonal to the electron path. The gamma factor $\gamma = E/m$ grows without bound with the electron energy E , so there is no upper limit on ω_c^{LI} .

Equation (3) is based on the electron trajectory in a given magnetic field, the radiation produced by a given current, and the relativistic relation between energy and velocity, all of which could be affected by Lorentz violation. We find that in the presence of Lorentz violation, equation (3) becomes:

$$\omega_c = \frac{3eBm\gamma(E)}{2m} \gamma^2(E) \quad (4)$$

where $\gamma(E) = \gamma(v(E))$. (For derivations of equation (4) and the constraint (equation (5)) below, see the Methods section.) If η is negative, electrons have a maximal velocity less than c , hence $\gamma(E)$ is a bounded function of E . This implies that there is a maximum achievable value of the cut-off frequency as given by equation (4), which we denote by ω_c^{max} . The value of the energy that yields this cut-off frequency is higher than the Lorentz-invariant value.

The rapid decrease in amplitude of synchrotron emission at frequencies larger than ω_c implies that most of the flux at a given frequency in a synchrotron spectrum is due to electrons whose ω_c is above that frequency. Thus ω_c^{max} must be greater than the maximum observed synchrotron emission frequency ω_{obs} . This yields the constraint:

$$\eta > -\frac{M}{m} \left(\frac{0.34eB}{m\omega_{\text{obs}}} \right)^{3/2} \quad (5)$$

The strongest synchrotron constraint comes from the system that minimizes the ratio B/ω_{obs} , which turns out to be the Crab nebula. The Crab nebula, a supernova remnant (SNR), is a bright source of radio, optical, X-ray and γ -ray emission, which exhibits a broad spectrum characterized by two marked humps. This spectrum is consistently explained by a combination of electron synchrotron emission and inverse Compton scattering of ambient photons by high energy electrons^{23,24}. In fact, no other model for the emission is under consideration, other than for producing some of the highest energy photons. For our constraint we assume that this

standard SNR model is correct.

The Crab synchrotron emission has been observed to extend at least up to energies of about 100 MeV (refs 23, 24), just before the inverse Compton hump begins to contribute to the spectrum. The magnetic field in the emission region has been estimated by several methods which agree on a value between 0.15 and 0.6 mG (see, for example, ref. 25 and references therein.) Two of these methods, radio synchrotron emission and equipartition of energy, are insensitive to Planck suppressed Lorentz violation, hence we are justified in adopting a value of this order for the purpose of constraining Lorentz violation. We shall use the largest value 0.6 mG for B since it yields the weakest constraint (equation (5) shows how the constraint scales with B).

Using the above values for the magnetic field and the highest synchrotron frequency, equation (5) yields the lower bound:

$$\eta > -7 \times 10^{-8} \quad (6)$$

This corresponds to:

$$E_{\text{QG}} = \frac{M}{|\eta|} > 10^{26} \text{ GeV} \quad (7)$$

in the phenomenological framework of ref. 6. Thus E_{QG} is constrained to be at least seven orders of magnitude larger than the Planck energy.

To complement the synchrotron constraint, which just bounds η from below, one can use the vacuum birefringence and Čerenkov constraints. Lack of observed vacuum birefringence bounds the difference in ξ for right and left circular polarized photons¹¹. Using the effective field theory result⁴ that left and right circular polarized photons have opposite values for ξ , this yields the constraint $|\xi| < 4 \times 10^{-4}$.

To bound η from above we can use the vacuum Čerenkov constraint. As noted by previous authors^{12,13,15,18}, modified dispersion can cause electrons to emit Čerenkov radiation in vacuum and rapidly lose energy. It was suggested in ref. 20 that a very strong Čerenkov constraint is obtained for positive η using the high energy electrons that produce the synchrotron radiation. But for positive η the energy required to produce a given cut-off synchrotron frequency goes down, as the electron group velocity can increase beyond the speed of light and $\gamma(E)$ can diverge at finite energy. Hence the electron energy may be much lower than it is in the Lorentz-invariant case. (Nevertheless a constraint can be obtained in this way; T. J., S. L., D. M. and F. Stecker, manuscript in preparation.) Thus we look to other high energy electrons whose existence is not in question.

The inverse Compton peak in the Crab spectrum contains energies up to 50 TeV. By energy conservation, this implies that electrons of at least 50 TeV propagate, hence the values of η and ξ must not allow these electrons to emit vacuum Čerenkov radiation. Within the region not ruled out by the birefringence bound, the results of refs 16–18 yield $\eta < 0.01$.

Finally, recall that we neglected ξ in obtaining the synchrotron constraint. Using equation (13) one can see that this is justified everywhere in the region not already excluded by the absence of vacuum Čerenkov radiation. The worst case would be at the smallest $|\eta|$. With η given by the lower bound (equation (6)), equation (13) shows that ξ can be neglected provided $|\xi| \lesssim 30$, which is a factor of nearly 10^3 larger than the vacuum birefringence limit.

Putting these observations together, we conclude that Lorentz violation suppressed by the ratio E/E_{Planck} and compatible with effective field theory is constrained in all directions of the ξ - η parameter space by much stronger than order unity bounds. The allowed region is a skinny rectangle, with bounds from vacuum birefringence above and below by $|\xi| < 4 \times 10^{-4}$, from Crab synchrotron radiation on the left by $\eta > -7 \times 10^{-8}$, and from vacuum Čerenkov radiation on the right by $\eta < 0.01$.

It is notable that, in such a short time after the recognition of the possibility of obtaining interesting bounds on Planck suppressed Lorentz violation, it has been so severely constrained. The constraint reported here on a possible maximal electron speed less than the speed of light is more than seven orders of magnitude better than previous limits. This limit is strong enough to conclude that quantum gravity scenarios postulating or implying this sort of Lorentz violation are not viable.

The type of Lorentz violation that we have considered also violates CPT symmetry⁴, the symmetry under combined charge conjugation, space inversion, and time reversal. The fact that it is ruled out might therefore be a consequence of CPT symmetry, rather than Lorentz symmetry. It thus makes sense to think seriously about constraining Lorentz violation suppressed by the second power, $(E/E_{\text{Planck}})^2$. Some interesting bounds already exist at this order^{16,26}, and other possibilities have been discussed^{16,27}. If it were eventually possible to achieve strong bounds at this order, then it would be reasonable to conclude that Lorentz symmetry is simply not violated, since there is no other symmetry that could protect against this sort of Lorentz violation. Alternatively, if the search for such constraints finds a Lorentz violation, that would open up a much-needed observational window on the difficult problem of quantum gravity. □

Methods

To determine the effects of Lorentz violation of equation (3), we follow the heuristic derivation of equation (3) given in ref. 22 assuming the framework of effective field theory. Without assuming Lorentz invariance, the purely kinematical result is:

$$\omega_c = \frac{3}{4} \frac{1}{R(E)\delta(E)} \frac{1}{c(\omega_c) - v(E)} \quad (8)$$

where $R(E)$ is the radius of curvature of the orbit, $\delta(E)$ is the opening angle for the forward-directed radiation pattern, and $c(\omega_c)$ and $v(E)$ are the group velocities of the radiation and electron, respectively. The solution of equation (8) for $\omega_c(E)$ determines the cut-off synchrotron frequency. In equation (8) we have used the fact that the electron and photon speeds are very close to the low energy speed of light c , which is set equal to unity. The numerical constant is chosen to yield the correct relativistic result equation (3), with $\delta(E) = \gamma^{-1}(E)$ (where $\gamma(E) = \gamma(v(E))$).

The radius $R(E)$ for a given energy is determined by the equation of motion of the electron in a magnetic field. All Lorentz violating terms in the equation of motion are suppressed by M^{-1} , and the leading high energy corrections come from modifications to the minimal coupling terms. To estimate the magnitude of the change, we use the dispersion relation as the hamiltonian. As usual, the minimal coupling is incorporated by replacing the momentum by $p - e\mathbf{A}$, where $\mathbf{A}$ is a vector potential for the magnetic field. This yields the equation of motion $\mathbf{a} = [1 + 3\eta E/2M](eE)\mathbf{v} \times \mathbf{B}$, where we have kept only the lowest order term in η and assumed relativistic energy $E \gg m$. Since $E \ll M$, the presence of the Lorentz violation makes very little difference to the orbital equation, hence we conclude that to a very good approximation the radius is related to the magnetic field and the energy of the electron by the standard formula $R(E) = E/eB$ (where again the speed of the electron has been set equal to unity).

The angle $\delta(E)$ scales in the Lorentz invariant case as $\gamma^{-1}(E)$. Since $R(E)$ and hence the charge current is nearly unaffected by the Lorentz violation, any significant deviation in $\delta(E)$ can only come from the modified response of the electromagnetic field to a given current. The amplitude of the modification is suppressed by at least one power of M , however, so dimensional analysis implies that it is suppressed by something of order $\xi\omega/M$. For the Crab synchrotron radiation used in our constraint ω is 100 MeV, which yields a suppression factor of $10^{-20}\xi$. Therefore the amplitude of the electromagnetic field modification is negligible in comparison to the zeroth order synchrotron radiation (which is still present), and hence $\delta(E)$ scales with $\gamma^{-1}(E)$ in the usual way.

Since the emitted photons have relatively low energy compared to the electrons, it turns out that ξ can be neglected in the relevant region of parameter space. Thus, since $v(E)$ is very close to c , the reciprocal of the difference of group velocities in the last term of equation (8) is well approximated by $2\gamma^2(E)$. This yields equation (4).

The maximum synchrotron frequency ω_c^{max} is obtained by maximizing ω_c (equation (4)) with respect to the electron energy, which amounts to maximizing $\gamma^3(E)/E$. The difference of group velocities is given by

$$c(\omega) - v(E) = \xi \frac{\omega}{M} + \frac{m^2}{2E^2} - \eta \frac{E}{M} \quad (9)$$

where we have used the dispersion relations (equations (1) and (2)) and dropped higher order terms. Dropping the ξ term this yields:

$$\gamma(E) \approx \left(\frac{m^2}{E^2} - 2\eta \frac{E}{M} \right)^{-1/2} \quad (10)$$

Using this expression we find:

$$\omega_c^{\text{max}} = 0.34 \frac{eB}{m} (-\eta m/M)^{-2/3} \quad (11)$$

which is attained at the energy $E_{\text{max}} = (-2m^2 M/5\eta)^{1/3} = 10(-\eta)^{-1/3}$ TeV. Since η is negative this is higher than the Lorentz invariant value that produces the same frequency, but only by a factor of order unity that works out to be $(9/5)^{3/4} \approx 1.55$. The energy of the electrons that produce the synchrotron radiation of frequency 100 MeV is 1,500 TeV in the Lorentz invariant case. If this were the maximum frequency in the Lorentz violating case, it would be produced by $\sim 2,300$ -TeV electrons.

We now analyse the question of neglecting ξ in evaluating the difference of group velocities (equation (9)). At the energy E_{max} the ratio of the ξ -dependent term to the other terms is given by:

$$\frac{|\xi|\omega}{(m^2 M/2E^2) - \eta E} = 3 \times 10^{-11} |\xi| (-\eta)^{-4/3} \quad (12)$$

Hence neglecting ξ is justified, provided that:

$$|\xi| \leq 10^{11} (-\eta)^{4/3} \quad (13)$$

Received 12 March; accepted 27 June 2003; doi:10.1038/nature01882.

- Gambini, R. & Pullin, J. Nonstandard optics from quantum spacetime. *Phys. Rev. D* **59**, 124021 (1999).
- Carroll, S. M., Harvey, J. A., Kostelecky, V. A., Lane, C. D. & Okamoto, T. Noncommutative field theory and Lorentz violation. *Phys. Rev. Lett.* **87**, 141601 (2001).
- Burgess, C. P., Cline, J., Filotas, E., Matias, J. & Moore, G. D. Loop-generated bounds on changes to the graviton dispersion relation. *J. High Energy Phys.* **0203**, 043 (2002).
- Myers, R. C. & Pospelov, M. Experimental challenges for quantum gravity. *Phys. Rev. Lett.* **90**, 211601 (2003).
- Pavlopoulos, T. G. Breakdown of Lorentz invariance. *Phys. Rev.* **159**, 1106–1110 (1967).
- Amelino-Camelia, G. *et al.* Tests of quantum gravity from observations of gamma-ray bursts. *Nature* **393**, 763–765 (1998).
- Schaefer, B. Severe limits on variations of the speed of light with frequency. *Phys. Rev. Lett.* **82**, 4964–4966 (1999).
- Biller, S. D. *et al.* Limits to quantum gravity effects from observations of TeV flares in active galaxies. *Phys. Rev. Lett.* **83**, 2108–2111 (1999).
- Kaaret, P. Pulsar radiation and quantum gravity. *Astron. Astrophys.* **345**, L32–L34 (1999).
- Norris, J. P., Bonnell, J. T., Marani, G. F. & Scargle, J. D. GLAST, GRBs, and quantum gravity. *Bull. Am. Astron. Soc.* **31**, 717 (1999).
- Gleiser, R. J. & Kozameh, C. N. Astrophysical limits on quantum gravity motivated birefringence. *Phys. Rev. D* **64**, 083007 (2001).
- Gonzalez-Mestres, L. Cosmological implications of a possible class of particles able to travel faster than light. *Nucl. Phys. Proc. Suppl.* **48**, 131–134 (1996).
- Coleman, S. & Glashow, S. L. High-energy tests of Lorentz invariance. *Phys. Rev.* **59**, 116008 (1999).
- Stecker, F. W. & Glashow, S. L. New tests of Lorentz invariance following from observations of the highest energy cosmic gamma rays. *Astropart. Phys.* **16**, 97–99 (2001).
- Jacobson, T., Liberati, S. & Mattingly, D. TeV astrophysics constraints on Planck scale Lorentz violation. *Phys. Rev. D* **66**, 081302 (2002).
- Jacobson, T., Liberati, S. & Mattingly, D. Threshold effects and Planck scale Lorentz violation: Combined constraints from high energy astrophysics. *Phys. Rev. D* **67**, 124011 (2003).
- Konopka, T. J. & Major, S. A. Observational limits on quantum geometry effects. *New J. Phys.* **4**, 57.1–57.8 (2002).
- Amelino-Camelia, G. Improved limit on quantum-spacetime modifications of Lorentz symmetry from observations of gamma-ray blazars. Preprint at (<http://arxiv.org/gr-qc/0212002>) (2002).
- Jacobson, T., Liberati, S. & Mattingly, D. Comments on 'Improved limit on quantum-spacetime modifications of Lorentz symmetry from observations of gamma-ray blazars'. Preprint at (<http://arxiv.org/gr-qc/0303001>) (2003).
- Stecker, F. W. Tests of quantum, gravity and large extra dimensions models using high energy gamma ray observations. Preprint at (<http://arxiv.org/astro-ph/0304527>) (2003).
- Gonzalez-Mestres, L. Lorentz symmetry violation and acceleration in relativistic shocks. *AIP Conf. Proc.* **558**, 874–877 (2001).
- Jackson, J. D. *Classical Electrodynamics*, 3rd edn, 671 (Wiley & Sons, New York, 1998).
- Atayan, A. M. & Aharonian, F. A. On the mechanisms of gamma radiation in the Crab Nebula. *Mon. Not. R. Astron. Soc.* **278**, 525–541 (1996).
- de Jager, O. C. *et al.* Gamma-ray observations of the Crab Nebula: A study of the synchro-Compton spectrum. *Astrophys. J.* **457**, 253–266 (1986).
- Hillas, A. H. *et al.* The Spectrum of TeV gamma rays from the Crab nebula. *Astrophys. J.* **503**, 744–759 (1998).
- Aloisio, R., Blasi, P., Ghia, P. L. & Grillo, A. F. Probing the structure of space-time with cosmic rays. *Phys. Rev. D* **62**, 053010 (2000).
- Amelino-Camelia, G. Proposal of a second generation of quantum-gravity-motivated Lorentz-symmetry tests: Sensitivity to effects suppressed quadratically by the Planck scale. Preprint at (<http://arxiv.org/gr-qc/0305057>) (2003).

Acknowledgements We thank F. A. Aharonian, G. E. Allen, G. Amelino-Camelia and F. Stecker for discussions. This work was supported in part by the NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.J. (jacobson@physics.umd.edu).

An intrinsic velocity-independent criterion for superfluid turbulence

A. P. Finne¹, T. Araki², R. Blaauwgeers^{1,3}, V. B. Eltsov^{1,4}, N. B. Kopnin^{1,5}, M. Krusius¹, L. Skrbek⁶, M. Tsubota² & G. E. Volovik^{1,5}

¹Low Temperature Laboratory, Helsinki University of Technology, PO Box 2200, FIN-02015 HUT, Finland

²Department of Physics, Osaka City University, Sumiyoshi-Ku, Osaka 558-8585, Japan

³Kamerlingh Onnes Laboratory, Leiden University, PO Box 9504, 2300 RA Leiden, The Netherlands

⁴Kapitza Institute for Physical Problems, and ⁵Landau Institute for Theoretical Physics, Kosygina 2, 119334 Moscow, Russia

⁶Joint Low Temperature Laboratory, Institute of Physics ASCR and Charles University, V Holešovičkách 2, 180 00 Prague, Czech Republic

Hydrodynamic flow in classical and quantum fluids can be either laminar or turbulent. Vorticity in turbulent flow is often modelled with vortex filaments. While this represents an idealization in classical fluids, vortices are topologically stable quantized objects in superfluids. Superfluid turbulence¹ is therefore thought to be important for the understanding of turbulence more generally. The fermionic ³He superfluids are attractive systems to study because their characteristics vary widely over the experimentally accessible temperature regime. Here we report nuclear magnetic resonance measurements and numerical simulations indicating the existence of sharp transition to turbulence in the B phase of superfluid ³He. Above 0.60T_c (where T_c is the transition temperature for superfluidity) the hydrodynamics are regular, while below this temperature we see turbulent behaviour. The transition is insensitive to the fluid velocity, in striking contrast to current textbook knowledge of turbulence². Rather, it is controlled by an intrinsic parameter of the superfluid: the mutual friction between the normal and superfluid components of the flow, which causes damping of the vortex motion.

In conventional liquids the vorticity $\omega = \nabla \times \mathbf{v}$ obeys the Navier-Stokes equation:

$$\frac{\partial \omega}{\partial t} = \nabla \times [\mathbf{v} \times \omega] + \nu \nabla^2 \omega \quad (1)$$

The interplay of the two terms on the right-hand side, the inertial first term and the viscous second term, governs the transition to turbulence (see, for example, ref. 2). It is determined by the external conditions through the Reynolds number, $Re = RU/\nu$, formed by the characteristic size R of the system, the flow velocity U , and the kinematic viscosity ν . At large $Re \gg 1$, the effect of the inertial term is dominating, and laminar flow becomes increasingly unstable. If vorticity is released into a metastable laminar state, a sudden transition to a chaotic flow of eddies occurs.

In superfluids turbulence acquires new features: a superfluid can be described as consisting of two inter-penetrating fractions, the frictionless (superfluid) and viscous (normal) components. If both fractions are moving together, as is the case behind a pulled grid in measurements on superfluid ⁴He-II (ref. 1), the turbulent state bears more resemblance to that of viscous liquids. A new class of superfluid turbulence becomes possible when the normal component is so viscous that it is essentially immobile and only the superfluid component is moving with respect to the boundaries. This is the case in ³He, which we consider here.

The vorticity of the superfluid component is quantized in terms of the circulation quantum $\kappa = 2\pi\hbar/M$ where M is the mass of a superfluid particle. The flow of the superfluid component can be characterized by the 'superfluid Reynolds number' $Re_s = RU/\kappa$. The Feynman criterion $Re_s \sim 1$ gives the velocity at which it becomes

energetically favourable to form a quantized vortex. If a large nucleation barrier exists, vortices are not created and the superfluid remains vortex-free, in our measurements up to $Re_s \approx 200$. Quantized vortices can then be injected, either by some extrinsic means or by increasing the velocity so that the nucleation barrier is overcome and spontaneous vortex formation occurs.

The limit of large $Re_s \gg 1$ is equivalent to a vanishing Planck's constant, $\kappa \propto \hbar \rightarrow 0$, so that the vorticity becomes a continuous variable, as in the classical case. However, the fluid still remains unconventional because of its two-fluid nature: in addition to the superflow, there is the normal component which is at rest in the container frame, $\mathbf{v}_n = 0$. Interactions between the superfluid vorticity and the normal component give rise to a mutual friction force on a unit volume of the superfluid:

$$\mathbf{f}_{mf} = -\alpha' \rho_s [\mathbf{v}_s \times \omega_s] + \alpha \rho_s [\hat{\omega}_s \times (\omega_s \times \mathbf{v}_s)]$$

which is described by the dimensionless temperature-dependent parameters α and α' . These correspond to the usual mutual friction parameters, if we are dealing with one vortex or locally polarized vorticity, or they may be renormalized in the general case. Here $\mathbf{v}_s$ and $\omega_s = \nabla \times \mathbf{v}_s$ are the superfluid velocity and vorticity, and $\hat{\omega}_s$ is the unit vector in the direction of ω_s . Inserting the mutual-friction force in the Euler equation one obtains the hydrodynamic equation for $Re_s \gg 1$ (ref. 3):

$$\frac{\partial \omega_s}{\partial t} = (1 - \alpha') \nabla \times [\mathbf{v}_s \times \omega_s] + \alpha \nabla \times [\hat{\omega}_s \times (\omega_s \times \mathbf{v}_s)] \quad (2)$$

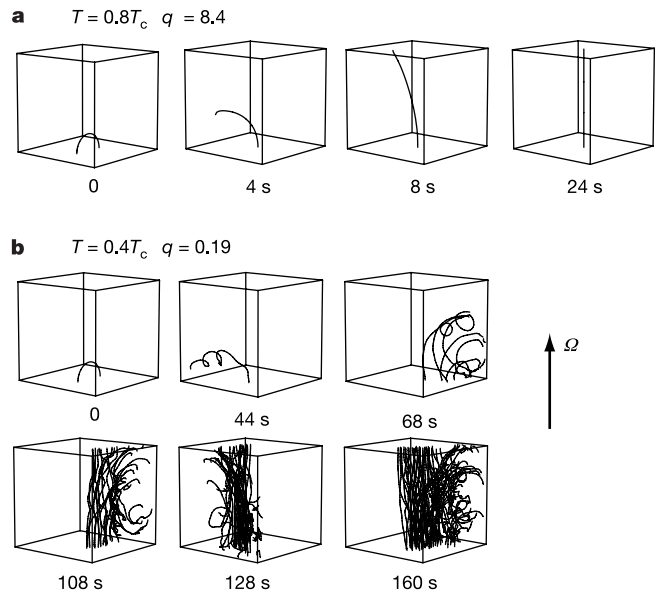


Figure 1 Summary of events at high and low temperatures. A vortex half-ring is injected into vortex-free superflow, generated with rotation around the vertical axis ($\Omega = 0.21 \text{ rad s}^{-1}$ or $Re_s \approx 30$). The plots represent snapshots from simulation calculations in the rotating frame and are consistent with the experimental observations. The only temperature dependence is contained in the mutual friction parameters α and α' which we take from the measurements in ref. 6. **a**, At high temperatures the loop grows monotonically into a single rectilinear vortex line, pulled by the Magnus force from the superflow. The motion is highly damped since the Magnus force is balanced by the mutual friction force. **b**, At low temperatures the loop moves with the circulating superflow. At 44 s the vortex is unstable with respect to loop formation and develops through reconnections into a tangle at the injection site. At 68 s the Magnus force from the vortex-free superflow extends the tangle along the vertical axis. At 108–160 s the final step is the reconnection of the tangle to rectilinear lines. The replacement of the large-scale superflow by the tangle and its later decay into rectilinear lines are observed as time-dependent frequency shifts in the NMR spectrum, while the spectrum in the final time-independent state yields the total number of rectilinear lines.

The energy dissipation is determined by the mutual friction damping α , while the reactive mutual-friction α' renormalizes the first term, the inertial term from equation (1).

As in classical fluids, the inertial term drives the instability and is responsible for the energy transfer to smaller length scales, whereas the dissipative term acts to stabilize laminar flow. The fundamental difference from conventional hydrodynamics is that the relative importance of these two competing terms is determined by an intrinsic parameter of the superfluid, $q = \alpha/(1 - \alpha')$. This is in contrast to classical liquids where this competition is governed via the Reynolds number by the extrinsic quantities R and U . In the case of a single vortex line the parameter q marks the crossover from the regime where Kelvin waves can propagate along a vortex ($q \lesssim 1$) to where they are overdamped ($q \gtrsim 1$) (ref. 4). In Fermi superfluids and superconductors with finite energy gap one has $q \approx (\omega_0 \tau)^{-1}$, where ω_0 is the spacing between the bound states of quasiparticles in the vortex core and τ^{-1} is their lifetime broadening, owing to scattering from the normal component⁵. In $^3\text{He-B}$, q approaches infinity at T_c and drops exponentially to zero, when cooled down to $T \rightarrow 0$ (ref. 6). Because in classical fluids such an evolution of q would correspond to scanning the Reynolds number from 0 to ∞ , one wonders how q influences the superfluid hydrodynamics.

This question can be answered with a straightforward measurement. In $^3\text{He-B}$ vortex-free superflow can be maintained indefinitely in uniform rotation up to high angular velocities Ω . Also other metastable states with rectilinear vortex lines, up to the equilibrium number $N_{\text{eq}} \propto \Omega$, are stable in time. By injecting vortex seed loops into the vortex-free state with different methods, we follow the

transient evolution of the loops until a final time-independent state is reached with N_f rectilinear vortex lines aligned along the rotation axis. Using non-invasive nuclear magnetic resonance (NMR) measurement, we record N_f as a function of temperature and rotation with a resolution better than ten vortex lines.

We find that above $0.60T_c$ an injected vortex loop expands into a single rectilinear vortex line; thus the number of vortices is here conserved, as shown in Fig. 1a. Below $0.60T_c$ the loops evolve at the injection site into a tangle, which spreads out axially filling the long sample cylinder, and then transforms to rectilinear lines with $N_f \approx N_{\text{eq}}$ (in qualitative agreement with the simulations in Fig. 1b). The difference in these two final states becomes obvious from a comparison of the NMR absorption spectra before and after injection: above $0.60T_c$ the changes are small since each injected loop gives rise to only one rectilinear vortex (Fig. 2a), whereas below $0.60T_c$ the spectra are totally different (Fig. 2b).

In the latter case the massive change in the NMR spectrum occurs over a transient period with characteristic time-dependent frequency shifts. These shifts as a function of time allow us to measure directly (1) the decay of the large-scale superflow, (2) the build-up of the high-density tangled vorticity, and (3) the subsequent formation of rectilinear lines (Fig. 2c). In the regular loop expansion process above $0.60T_c$ the long-lived transients are not present. Our 11-cm-long sample (with radius $R = 3$ mm) is monitored with two independent NMR detectors at both its ends⁷, to follow the evolution of the transients simultaneously at these two locations. The correlations in time prove that the tangle expands along the sample, persists over a characteristic transient period, and does not

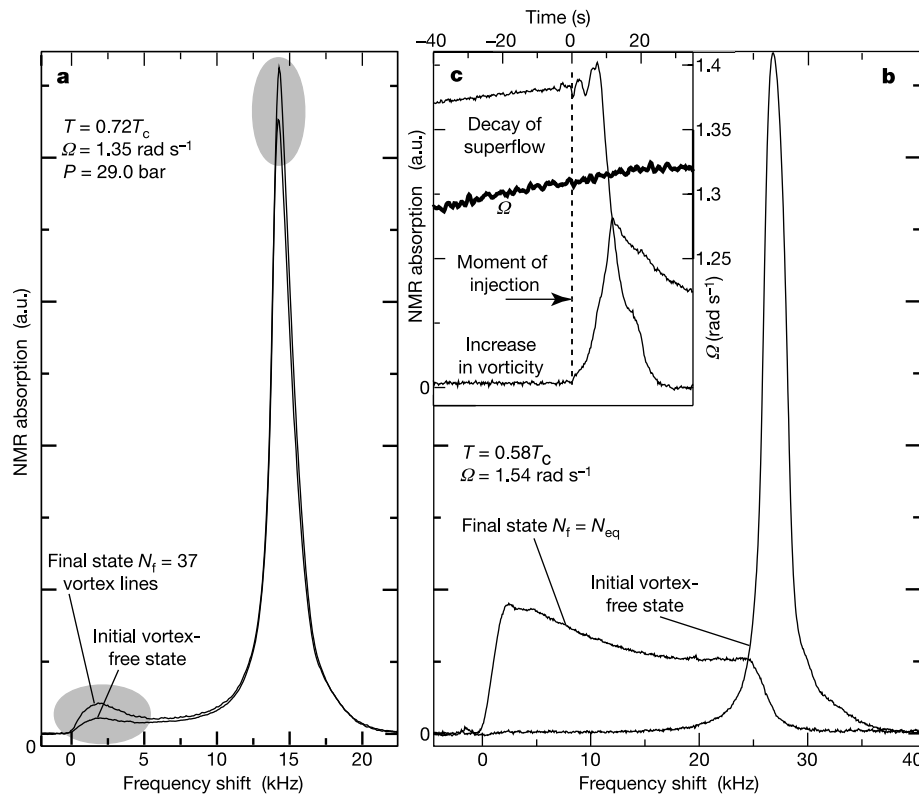


Figure 2 NMR absorption spectra before and after vortex-loop injection. **a**, Regular mutual-friction-damped loop expansion above $0.60T_c$ introduces only minor changes in the spectrum. The sharp peak at large frequency shift is caused by the large-scale superflow (in the rotating frame). When a few rectilinear vortex lines are formed, some absorption from the large peak is shifted into a tiny new peak at small frequency shift. The heights of both peaks change linearly with N_f (in the highlighted regions). **b**, Below $0.60T_c$ a radical restructuring of the spectrum occurs when the large-scale superflow is

replaced by an equilibrium array of rectilinear vortices. **c**, Peak heights of the two absorption maxima as a function of time during the transient turbulent period: (1) the sharp peak at large frequency shift decreases which monitors the decay of the large-scale superflow. (2) The tiny peak at small frequency shift first increases, which reflects the build up of the high-density vortex tangle. Subsequently it starts to decrease, which corresponds to the rarefaction and polarization of the tangle into an equilibrium array.

depend on the details of the injection.

In Fig. 3 we list the results on the vortex-line count N_f in the (Ω, T) plane. Regular high-temperature processes are marked with open symbols and turbulent low-temperature points with filled symbols. The transition between them at $0.60T_c$ is found to be surprisingly sharp, with $N_f \sim 1$ above and $N_f \approx 10^3$ below the transition. Only within a narrow interval of $0.05T_c$ width can either type of process occur randomly. Even in this overlap region intermediate line counts are rare, where one would find $1 \ll N_f \ll N_{eq}$, and thus no high-resolution classification is needed in Fig. 3. The central conclusion is that the transition at $0.60T_c$ occurs as a function of temperature, with little dependence on the maximum superflow velocity $U = \Omega R$ of the initial vortex-free state. At $0.60T_c$ the value of $q = \alpha/(1 - \alpha') \approx 1.3$ (ref. 6).

Figure 1 shows that the measured features are qualitatively confirmed by our simulation calculations. These are started from an initial state with a vortex half ring (diameter 3.5 mm) in the centre of a cubic sample container (10 mm wide). At high temperatures the high damping α causes the seed loop to expand, mainly by motion in the radial and axial directions, into a single rectilinear vortex line in the centre of the sample. At low temperatures the seed loop travels azimuthally and helical Kelvin waves start to expand on such sections where the vortex is oriented along the flow. Here $q < 1$, Kelvin waves are only lightly damped and grow to loops which via reconnections lead to tangle formation (ref. 8). Thus a

vortex network develops already near the injection site and then extends axially. Through reconnections a bundle of lines starts to form which precesses initially around the $\hat{\Omega}$ axis. More reconnections of remaining loops continue adding lines to the bundle, thereby reducing the superflow around it, until the equilibrium number of rectilinear lines is reached. The complete calculation is, so far, too time-consuming.

Both the results from the numerical simulations in Fig. 1 and from the measurements in Fig. 3 support the conclusion from equation (2) that the transition to turbulence is determined by the intrinsic parameter $q = \alpha/(1 - \alpha')$: there exists a critical value $q_c \approx 1$ which separates the mutual-friction-dominated regime at $q > q_c$ from the inertia-dominated turbulent regime at $q < q_c$. With $q > q_c$, the vortex-loop expansion is a regular stable process even in high-velocity superflow with $Re_s \gg 1$. With $q < q_c$, a multiplication process is switched on, the vortex lines create loops, become entangled, and reconnect forming new loops. Thus we find q_c to be a fundamental universal number which places the upper bound to the stability of superfluid turbulence in the limit $Re_s \gg 1$.

This result resolves the existing conflict between high- and low-temperature measurements in $^3\text{He-B}$: at high temperatures a vortex loop injected into rotating superflow expands to a single rectilinear vortex^{7,9,10}, while at low temperatures a tangle of vorticity can be generated in a quiescent bath with a vibrating wire¹¹. But why has

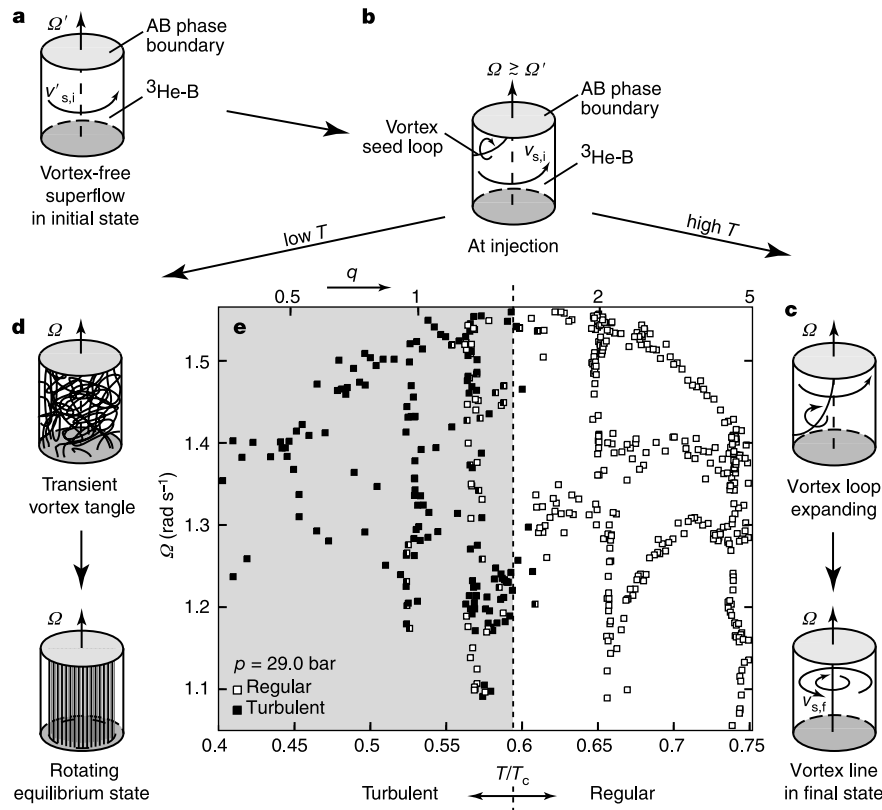


Figure 3 Principle of measurement and phase diagram of turbulent superflow in $^3\text{He-B}$. **a**, The initial state is vortex-free superflow in rotation at Ω , where the normal component is stationary and the superfluid component flows in the rotating frame. **b**, A few (ΔN) vortex loops are injected and, after a transient period of loop expansion, the number of rectilinear vortex lines N_f in the final steady state is measured. It is found to fall in one of two categories. **c**, $N_f = \Delta N$, regular mutual-friction-damped loop expansion. **d**, $\Delta N \ll N_f \lesssim N_{eq}$, turbulent loop expansion. This process leads to a total removal of the macroscopic vortex-free superflow as the superfluid component is forced into solid-body-like rotation (on average) by the formation of a vortex array with the equilibrium

number of rectilinear lines, $N_{eq} \approx \pi R^2 2\Omega/\kappa \approx 10^3$. **e**, Phase diagram of measured events: each data point monitors the outcome from the injection event in the final steady state as a function of rotation and temperature. Regular loop expansion is marked with open symbols and turbulent with filled symbols. An abrupt transition at $0.60T_c$, independently of Ω , is found to divide vertically the (Ω, T) plane between the two types of vortex dynamics. Only within a narrow interval $\Delta T \approx 0.05T_c$ around $0.60T_c$ can intermediate behaviour be observed. The injection is here initiated using the shear-flow instability of the interface between $^3\text{He-A}$ and $^3\text{He-B}$ (ref. 7).

such a transition not been observed previously in $^4\text{He-II}$? A number of reasons can be listed: (1) in ^3He the viscosity of the normal fraction is four orders of magnitude higher. Thus there is no ambiguity about the state of the normal component, unlike in the case of $^4\text{He-II}$. (2) The vortex-core radius in $^3\text{He-B}$ exceeds by two orders of magnitude that in $^4\text{He-II}$, where it is only of atomic size. Therefore vortex pinning and remanence at solid walls can often be neglected in $^3\text{He-B}$ and metastable states with high-velocity superflow become possible. (3) The most important difference lies in the value of mutual friction. In $^4\text{He-II}$ the dissipative part α is small and influences vortex motion relatively little. Only within a narrow temperature interval $\Delta T/T_\lambda \approx 2 \times 10^{-3}$ from T_λ one finds $q > 1$ and expects different behaviour. No conclusive measurements exist from so close to T_λ .

To conclude we may now predict that in helium superfluids the transition at $q_c \approx 1$ lies close to T_λ in $^4\text{He-II}$, in the middle of the experimentally accessible temperature range in $^3\text{He-B}$, and below that at extremely low temperatures in $^3\text{He-A}$. This prediction is consistent with presently available experimental information. Perhaps in other hydrodynamic systems, which consist of two or more components, superfluid or viscous, the stability of turbulence might also be governed by an intrinsic parameter, in addition to the Reynolds number. □

Methods

Vortex injection

We use four different injection methods: (1) vortex loops are created at a rough spot on the cylindrical sample boundary by increasing Ω to the critical value Ω_c (ref. 9). (2) In neutron irradiation, vortices are created via the heat release from neutron capture reactions¹⁰. (3) Vortex loops may leak through the small orifice which connects the NMR sample to the rest of the liquid ^3He volume¹². (4) The most efficient method is to use an instability of the phase boundary between the $^3\text{He-A}$ and $^3\text{He-B}$ phases in rotation⁷. The AB boundary can be stabilized to a fixed location in a suitable magnetic field gradient. This technique has been employed to collect the data in Fig. 3, to fill the (Ω, T) plane with a dense grid of measurements. The instability of the AB interface produces a small random number ($\Delta N \approx 10$) of B-phase vortex loops of which one end sticks out of the AB interface while the other end travels along the cylindrical sample boundary during the expansion process into rectilinear vortices⁷. The four injection processes have different temperature-dependent critical velocities $\Omega_c(T)$, which are continuous at $0.60T_c$ and consistently display the transition from regular to turbulent dynamics at $0.60T_c$.

NMR signatures from turbulence

The lowest measured critical velocities are about 0.5 rad s^{-1} at $T \leq 0.50T_c$. Even at these velocities ($\text{Re}_\lambda \leq 70$) the injection leads only to turbulent loop expansion. The axial expansion of the vorticity along the Ω axis between the two signal coils is found to be controlled by the mutual-friction damping α : the time of expansion τ across a vertical d agrees with the expression $\tau = d/(\alpha\Omega R)$. The extracted $\alpha(T)$ reproduces the result of ref. 6, is continuous across the transition at $0.60T_c$, and is thus indifferent to whether the vorticity moves as noninteracting loops or as a network of loops. Below the transition at $0.60T_c$ the vortex tangle is observed to decay with a time constant of around 30 s which decreases towards lower temperatures. This we interpret to indicate that reconnection processes are primarily responsible for the decay rather than the mutual-friction dominated expansion of individual loops.

Simulation

We use the vortex filament model^{13,14} in the rotating frame¹⁵. A vortex is represented in parametric form by $\mathbf{s} = \mathbf{s}(\xi, t)$, where $\mathbf{s}$ refers to a point on the filament and ξ is the arc length along it. The spatial and time evolutions are integrated rigorously using the Biot–Savart law. The vortex velocity $\dot{\mathbf{s}}$ is calculated from the dynamic equation $\dot{\mathbf{s}} = \mathbf{v}_{sl} + \alpha' \mathbf{s}' \times [\mathbf{s}' \times \mathbf{v}_{sl}] - \alpha[\mathbf{s}' \times \mathbf{v}_{sl}]$ (recall that $\mathbf{v}_n = 0$), where the local superfluid velocity $\mathbf{v}_{sl}$ includes all contributions to the superflow at $\mathbf{s}(\xi, t)$. As boundary conditions we use smooth solid-walls with image vortices.

Analytic approach

Equation (2) is a first-order differential equation in contrast to the second order Navier–Stokes equation (1) of conventional hydrodynamics. Our result suggests a new approach to classical turbulence: when the vorticity is modelled with stable vortex filaments, their dynamics can be derived from a first-order equation by introducing an effective vortex-filament viscosity.

Received 18 March; accepted 1 July 2003; doi:10.1038/nature01880.

- Vinen, W. F. & Niemela, J. J. Quantum turbulence. *J. Low-Temp. Phys.* **128**, 167–231 (2002).
- McComb, W. D. *The Physics of Fluid Turbulence* 2 (Clarendon, Oxford, 1990).
- Sonin, E. B. Vortex oscillations and hydrodynamics of rotating superfluids. *Rev. Mod. Phys.* **59**, 87–155 (1987).
- Barenghi, C. F. *et al.* Thermal excitation of waves on quantized vortices. *Phys. Fluids* **28**, 498–504 (1985).

- Kopnin, N. B. *Theory of Nonequilibrium Superconductivity* 271 (Clarendon, Oxford, 2001).
- Bevan, T. D. C. *et al.* Vortex mutual friction in superfluid ^3He . *J. Low-Temp. Phys.* **109**, 423–459 (1997).
- Blaauwgeers, R. *et al.* Shear flow and Kelvin–Helmholtz instability in superfluids. *Phys. Rev. Lett.* **89**, 155301 (2002).
- Ostermeier, R. M. & Glaberson, W. I. Instability of vortex lines in the presence of axial normal fluid flow. *J. Low-Temp. Phys.* **21**, 191–196 (1975).
- Ruutu, V. M. H. *et al.* Intrinsic and extrinsic mechanisms of vortex formation in superfluid ^3He . *J. Low-Temp. Phys.* **107**, 93–164 (1997).
- Ruutu, V. M. H. *et al.* Vortex formation in neutron-irradiated superfluid ^3He as an analogue of cosmological defect formation. *Nature* **382**, 334–336 (1996).
- Fisher, S. N. *et al.* Generation and detection of quantum turbulence in superfluid $^3\text{He-B}$. *Phys. Rev. Lett.* **86**, 244–247 (2001).
- Skrbek, L. *et al.* Vortex flow in rotating superfluid $^3\text{He-B}$. *Physica B* **329–333**, 106–107 (2003).
- Schwarz, K. W. Three-dimensional vortex dynamics in superfluid ^4He : Homogenous superfluid turbulence. *Phys. Rev. B* **38**, 2398–2417 (1988).
- Tsubota, M. *et al.* Dynamics of vortex tangle without mutual friction in superfluid ^4He . *Phys. Rev. B* **62**, 11751–11762 (2000).
- Tsubota, M. *et al.* Rotating superfluid turbulence. *Phys. Rev. Lett.* **90**, 205301 (2003).

Acknowledgements This work was supported in part by the EU-IHP ULTI-3, ESF-COSLAB, and ESF-VORTEX programmes. N.B.K. and G.E.V. are grateful to the Russian Foundation for Basic Research and L.S. to the Grant Agency of the Czech Republic. We thank C.F. Barenghi, P.V.E. McClintock and W.F. Vinen for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.E. (ve@boojuum.hut.fi).

Water-driven structure transformation in nanoparticles at room temperature

Hengzhong Zhang*, Benjamin Gilbert*, Feng Huang & Jillian F. Banfield

Department of Earth and Planetary Sciences, University of California, Berkeley, California 94720, USA

* These authors contributed equally to this work

The thermodynamic behaviour of small particles differs from that of the bulk material by the free energy term γA —the product of the surface (or interfacial) free energy and the surface (or interfacial) area. When the surfaces of polymorphs of the same material possess different interfacial free energies, a change in phase stability can occur with decreasing particle size^{1,2}. Here we describe a nanoparticle system that undergoes structural changes in response to changes in the surface environment rather than particle size. ZnS nanoparticles (average diameter 3 nm) were synthesized in methanol and found to exhibit a reversible structural transformation accompanying methanol desorption, indicating that the particles readily adopt minimum energy structural configurations^{3,4}. The binding of water to the as-formed particles at room temperature leads to a dramatic structural modification, significantly reducing distortions of the surface and interior to generate a structure close to that of sphalerite (tetrahedrally coordinated cubic ZnS). These findings suggest a route for post-synthesis control of nanoparticle structure and the potential use of the nanoparticle structural state as an environmental sensor. Furthermore, the results imply that the structure and reactivity of nanoparticles at planetary surfaces, in interplanetary dust⁵ and in the biosphere^{6,7}, will depend on both particle size and the nature of the surrounding molecules.

ZnS belongs to a class of important II–VI semiconductors that includes CdS and CdSe. Cadmium chalcogenides are often synthesized as nanoparticles for studies of size-dependent phenomena,

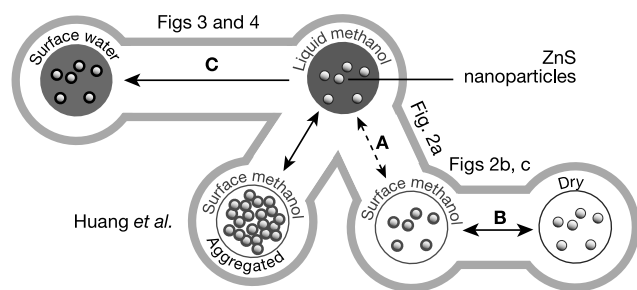


Figure 1 Diagram showing the experiments (A, B and C) performed in this and related work^{3,4}. The dashed double-headed arrow indicates a reversible step with no structural change; solid double-headed arrows indicate cases where nanoparticle's surface environment causes reversible structural modification. The single-headed arrow indicates that the water-driven transformation has not been reversed, probably owing to strong chemisorption of the water at the nanoparticle surface.

such as quantum confinement⁸. In CdSe, changes in size correlate with changes in colour that span the visible (vis.) spectrum^{9,10}, while confinement effects in ZnS appear in the ultraviolet (UV) region¹¹. Here we integrate structure modelling and materials characterization methods to demonstrate the ways in which ZnS nanoparticle structure responds to changes in the surface environment. The experimental framework is shown in Fig. 1.

Nanocrystalline ZnS was synthesized in anhydrous methanol without a surfactant. The average particle diameter determined from the edge position of the UV-vis. absorption threshold is 2.8 ± 0.2 nm, and when determined by transmission electron microscopy (TEM) is 3.0 ± 0.3 nm (Supplementary Figs 1 and 2). If the nanoparticle size is determined by X-ray diffraction (XRD) using the Scherrer equation¹², which only considers size contributions to peak broadening, the estimated diameter is 1.4 nm. The discrepancy implies that significant strain is present within the nanoparticles.

We demonstrated that the ZnS nanoparticles are not trapped in a metastable initial state but can undergo reversible structural transformations at room temperature. For example, Huang *et al.*^{3,4} showed that when particles are ultrasonically disaggregated at room temperature and allowed to reaggregate, a reversible structural transformation occurs. In addition, we tested for reversible

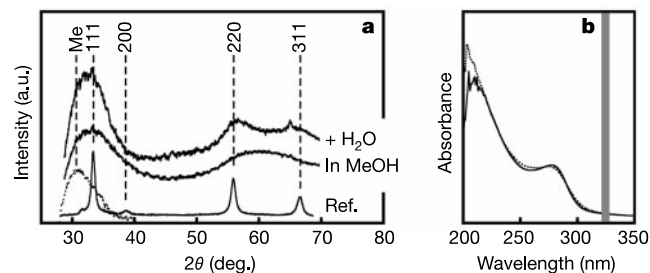


Figure 3 The effect of water binding on the structure and size of uncapped ZnS nanoparticles in methanol. See step C in Fig. 1. **a**, Conventional XRD reveals a dramatic increase in crystallinity after addition of water, as the 220 and 311 reflections become resolved. Low angle diffraction from the methanol solvent ("Me") interferes with the 111 peak; 'Ref.' indicates 20 nm ZnS (mostly sphalerite, commercial, Aldrich). **b**, UV-vis. absorption spectroscopy shows no change in the onset of absorption, indicating that no coarsening occurs due to water addition. Solid line, sample in methanol; dotted line, following addition of water. The vertical bar indicates the absorption threshold position for bulk sphalerite ZnS (see Supplementary Fig. 1).

structural transformations accompanying removal and resorption of methanol.

Methanol-suspended ZnS nanoparticles were transferred to vacuum at room temperature. Liquid methanol rapidly evaporated, but XRD of the ZnS nanoparticles showed no structural change (Fig. 2a). Thermal desorption measurements revealed that methanol was retained on the nanoparticle surface and was evolved above 50 °C (Supplementary Information). Extended X-ray absorption fine structure (EXAFS) measurements showed that after thermal desorption of methanol at 50 °C the nanoparticles are structurally distinct from those in the presence of methanol. EXAFS data indicated that the nanoparticles transform from a distorted four-coordinate structure (in methanol) to an unidentified phase (dry) (Supplementary Table 1 and Supplementary Fig. 3). The real-space EXAFS correlation functions indicated that the original structure was regained after readdition of methanol (Fig. 2b, c). Thus, the experiment demonstrates a reversible structural transformation associated with methanol desorption and replacement.

To study the effect of water addition, 50 μ l water per ml methanol was added to the nanocrystalline ZnS in liquid methanol, and the sample characterized *in situ* (no drying). Figure 3a and b shows XRD patterns and UV-vis. spectra of the nanocrystalline ZnS in methanol without water, and 24 h after addition of water. The XRD

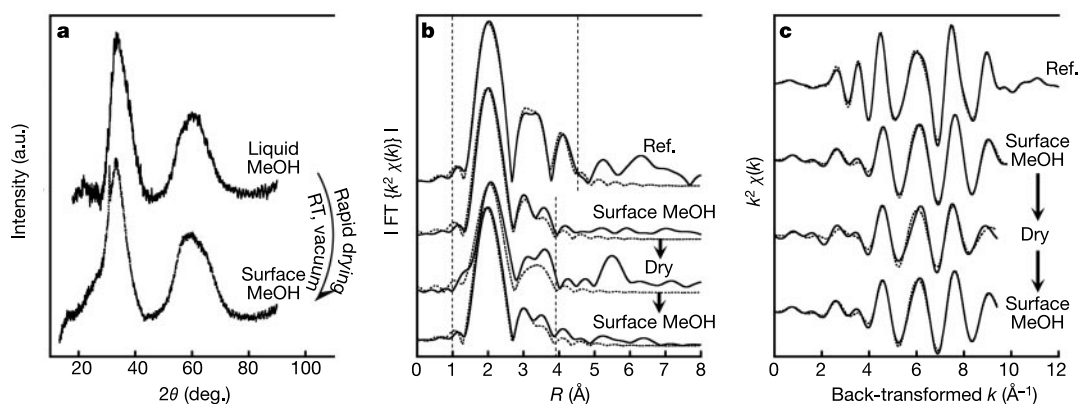


Figure 2 A reversible structural change associated with methanol desorption. **a**, *In situ* X-ray diffraction of ZnS nanoparticles shows that rapid pre-drying (step A in Fig. 1) does not induce structure change. When surface methanol (MeOH) is removed by thermal desorption at 50 °C (step B, Fig. 1) the EXAFS data (**b**, Fourier transforms (FT) of k^2 weighted $\chi(k)$ EXAFS functions, **c**, back transformed data) indicate a structural

modification that can be reversed by readdition of methanol. EXAFS spectra labelled "Ref." are from bulk ZnS (sphalerite). The vertical dashed lines are the real-space fitting ranges (three-shells for the reference, two-shells for nanoparticles). Data and fits are given by solid and dotted curves, respectively.

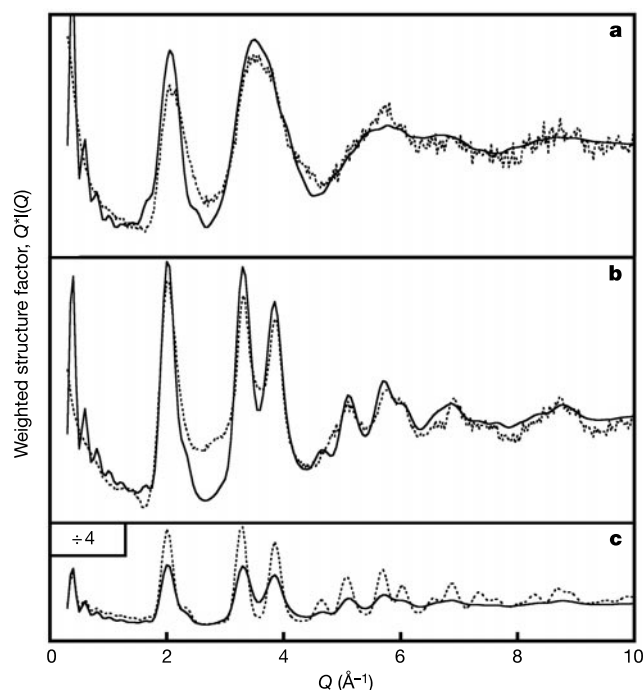


Figure 4 Wide-angle X-ray scattering (WAXS) observation of water binding. Experimental (dotted lines) WAXS Q -weighted structure factors of ZnS nanoparticles in methanol **a**, before and **b**, after water binding (step C in Fig. 1). Superimposed (solid lines) are the patterns predicted by molecular dynamics (MD) relaxation of a 3 nm ZnS particle **a**, without and **b**, with surface bound water. The methanol background has been subtracted from the experimental data. **c**, The WAXS pattern from the 3 nm MD simulation in the presence of water (solid line) is compared with the pattern of an unrelaxed 3 nm spherical fragment of the sphalerite structure (dotted line).

patterns clearly demonstrate structural change. However, the UV-vis. absorption threshold position is unchanged, indicating that no significant size change occurs following water addition (Fig. 3b). TEM imaging of nanoparticles without water, and following the addition of water, also revealed no size change (Supplementary Fig. 2). Peak positions in the XRD spectrum following water addition suggest that the disordered structure undergoes a transformation to more crystalline sphalerite-like structure.

The water-driven structure transformation in ZnS was further studied with synchrotron-based wide angle X-ray scattering (WAXS) to achieve higher quality *in situ* diffraction measurements to high $Q = (4\pi\sin\theta)/\lambda$ with accurate background subtraction, and to obtain the real-space pair correlation function $G(r)$ (Supplementary Fig. 4). From the conventional XRD data in Fig. 3 and the WAXS data in Fig. 4 we conclude that nanocrystalline ZnS in methanol undergoes a room temperature structural modification, best characterized as a disorder-to-order transition, following water binding. Identical structural behaviour was observed with *in situ* X-ray absorption near edge structure (XANES) and EXAFS spectroscopies (Supplementary Figs 5a, b).

Further insight into the initial and final structures and into the transition itself are obtained by molecular dynamics (MD) simulations of model particles of different sizes (2.0, 2.5, 3.0, 4.0 and 5.0 nm), from different initial structures (tetrahedrally coordinated hexagonal (wurtzite) or cubic (sphalerite) ZnS), and with different water coverages (1, 3 and 5 H_2O molecules per nm^2 ZnS surface). The MD simulations were performed with a time step of 5×10^{-4} ps (up to 100 ps or more MD time) until a full relaxation of the initial configuration was achieved without loss of the original phase (see Methods).

We calculated the theoretical WAXS pattern associated with each final MD structure. The experimental trend towards increased crystallinity following water binding was reproduced by the MD series. The closest agreement with data from nanoparticles in

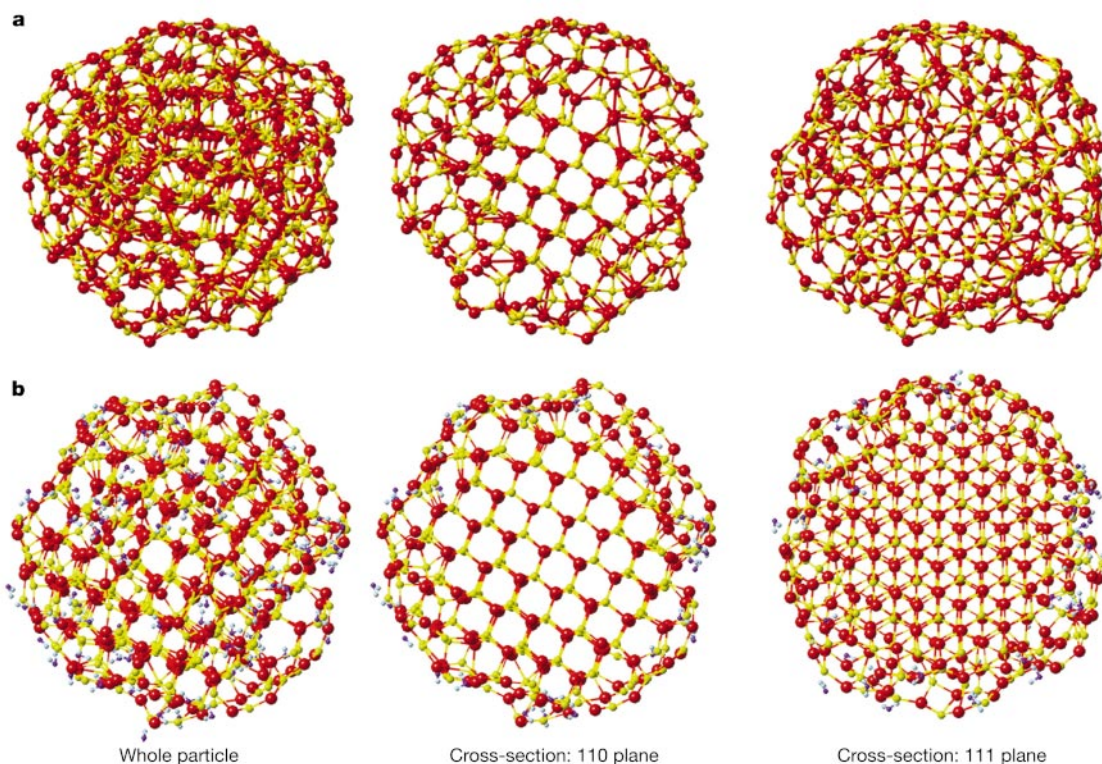


Figure 5 Molecular dynamics predictions of the structure of a 3 nm ZnS nanoparticle. **a**, Without surface-bound water, and **b**, with surface-bound water. S atoms yellow, Zn red, O blue, H light blue. Central cross-sections through the particles give a clearer picture

of internal structure. Regular interior (110) and (111) planes are evident in both structures, but in the absence of water ligands the outer shell is severely distorted.

methanol was obtained from MD relaxation of an unpassivated 3 nm sphalerite nanocrystal. MD relaxation of 2.5 and 3 nm ZnS nanocrystals (with water coverage at 3 H₂O molecules per nm²) both gave close agreement with data in the presence of water. In each case, the MD particle size is close to the actual nanoparticle size. MD simulations of particles that were not close in size to the experimental nanoparticles gave poor fits to the data. The corresponding theoretical WAXS patterns are displayed in Fig. 4, and views of the nanoparticle structures are given in Fig. 5.

The experimental X-ray scattering data and MD calculations indicate a highly disordered structure within the ZnS nanoparticles in methanol. However, cross-section views through the MD structure (Fig. 5) reveal a crystalline central core, indicating that the surfaces of unpassivated nanoparticles undergo reconstruction that is more extreme and penetrates deeper than that documented previously at macroscopic surfaces¹³ and step edges¹⁴. Simulations suggest that distortions due to the cumulative effects of bounding surfaces in three dimensions affect at least four atomic layers, approximately 0.8 nm. Analogous distortion has been inferred from studies of cold noble gas clusters^{15,16}, but not in covalently bonded solids.

We also attempted to model the WAXS data using fragments of perfect sphalerite or wurtzite that contained stacking faults and thermal and structural disorder. None of these models generated fits to the data that were as good as those from MD simulations. This finding is significant, as perfect structures are frequently the starting point for calculations of nanoparticle properties^{17–19}. The existence of surface relaxation and reconstruction has been proposed on the basis of theoretical^{20,21} and experimental studies^{22,23}, but no previous study has obtained agreement between structural data and simulations that include interfacial interactions.

The MD simulations show that interactions between water and ZnS decrease the interfacial energy. The resulting increase in crystallinity propagates through the nanoparticle, driven in the simulation by kinetic energy obtained from the isothermal environment. The low water coverage required to increase the crystallinity in the MD simulations indicates a strong water interaction with the surface, consistent with the surface chemistry of hydrated ZnS (ref. 24). The enthalpy change associated with water adsorption was estimated from the MD simulations to be ~500 kJ per mol H₂O at 3 H₂O per nm² ZnS surface coverage, but includes both the heat of adsorption plus structural stabilization energy. By comparison, the calorimetrically determined heat of water adsorption for nanocrystalline alumina is 250–300 kJ per mol H₂O at 3 H₂O per nm² Al₂O₃ surface coverage¹. The MD prediction for nanocrystalline ZnS therefore indicates both strong interaction between water and the surface, and a very large stabilization effect. The MD simulations predict that the polar water molecules orient to permit hydrogen and oxygen bonding to the terminating S and Zn ions. Surface interactions with methanol (not included here) are presumably weaker owing to its lower molecular polarity.

Although a more crystalline structure is obtained when water interacts with the nanoparticle, inspection of the surface structure shows residual surface strain, in agreement with prior EXAFS analyses of covalently capped CdTe nanoparticles²⁵. Comparison of theoretical WAXS patterns for relaxed and unrelaxed 3 nm particles (Fig. 4c) shows that the structure also retains disorder relative to perfect sphalerite. □

Methods

Nanoparticle synthesis and transformation

3 nm ZnS nanoparticles were synthesized in anhydrous methanol by dropwise addition of 1 M ZnCl₂ to 1 M Na₂S under a N₂ atmosphere at room temperature. The resulting suspension was washed three times and stored in methanol. Two 1.9 ml aliquots of the homogenous product were removed, and 100 µl H₂O or methanol added. Size and structure analyses were performed 24 h later on both samples.

X-ray diffraction

A Bruker D8 Discover GADDS diffractometer operated in the reflective geometry with Co

K_α radiation (λ = 1.790 Å) was used to measure dilute suspensions of ZnS nanoparticles in methanol in a low background quartz wet-cell with a thin Kapton window in a flat plate geometry.

UV-vis. spectroscopy

UV-vis. absorption spectra were acquired with an Agilent 8453 spectrophotometer, from synthesized and transformed ZnS nanoparticles in methanol following 4:1 dilution.

WAXS

WAXS was performed on the 11-ID-C beamline of the Advanced Photon Source, Argonne, Illinois, with X-ray wavelength λ = 0.10759 Å. Suspensions of ZnS nanoparticles were placed in cylindrical plastic containers (diameter 7 mm, beam path 6.5 mm) with Kapton entrance and exit windows. No signal from water (g(OH) or g(OO)) was observed in the real space or reciprocal WAXS patterns²⁶. Theoretical calculations of WAXS intensity were derived from the atomic coordinates of models of the nanocrystals using the Debye equation, incorporating a fitted Debye Waller factor²⁷:

$$I(Q) = \sum_i \sum_j f_i f_j \frac{\sin(Qr_{ij})}{Qr_{ij}} \exp[-2M]$$

MD simulations

Zinc sulphide was described using a shell model, designed to account for molecular polarity, with a Buckingham form of the pair-wise interatomic potential functions, plus an angle-bending form of the three-body interaction between nearest neighbour S–Zn–S atoms²⁸. Using this description of ZnS, at 300 K the predicted lattice parameters of the bulk sphalerite, wurtzite and the rock-salt phases of ZnS are, respectively, no more than 0.1%, 2.8% and 0.3% different from the literature values. The predicted transition pressure from sphalerite to rock-salt phase at 300 K is 12 GPa, consistent with ref. 29. In the evaluation of three water models, the shell model³⁰ showed best compatibility with the ZnS potentials, and was selected over the simple point charge SPCE³¹ and central force CF³² models. The short range Zn–O and S–H interactions also assume a Buckingham potential function, the former taken from ref. 33, the latter obtained from a fit to first principles data for the ground state H–HS potential well in gas phase H₂S (ref. 34).

$$u_{S-H}(\text{short-range}) = 15,644.3 \exp\left(-\frac{R_{S-H}}{0.1565}\right) - \frac{21.289}{R_{S-H}^6},$$

where *u* is in eV and *R* in Å. The MD simulations were run in a canonical ensemble at 300 K using a Nose–Hoover algorithm, or (optionally, at large times) a gaussian algorithm, with a time step of 0.5 fs for no less than 100 ps.

Received 23 April; accepted 12 June 2003; doi:10.1038/nature01845.

- McHale, J. M., Auroux, A., Perotta, A. J. & Navrotsky, A. Surface energies and thermodynamic phase stability on nanocrystalline aluminas. *Science* **277**, 788–791 (1997).
- Zhang, H. & Banfield, J. F. Thermodynamic analysis of phase stability of nanocrystalline titania. *J. Mater. Chem.* **8**, 2073–2076 (1998).
- Huang, F., Zhang, H., Gilbert, B. & Banfield, J. F. Surface state controlled nanocrystalline ZnS structure transformation. *Abstr. Pap. Am. Chem. Soc.* **225**, 431-Phys (2003).
- Huang, F., Zhang, H., Gilbert, B. & Banfield, J. F. Reversible, surface-controlled structure transformation in nanoparticles induced by aggregation-disaggregation. *Phys. Rev. Lett.* (submitted).
- Li, A. & Draine, B. T. Are silicon nanoparticles an interstellar dust component? *Astrophys. J.* **564**, 803–812 (2002).
- Labrenz, M. *et al.* Formation of sphalerite (ZnS) deposits in natural biofilms of sulphate-reducing bacteria. *Science* **270**, 1744–1747 (2000).
- Banfield, J. F. & Navrotsky, A. (eds) *Nanoparticles and the Environment* (Reviews in Mineralogy and Geochemistry, Vol. 44, Mineralogical Society of America, Washington, 2001).
- Alivisatos, A. P. Perspectives on the physical chemistry of semiconductor nanocrystals. *J. Phys. Chem.* **100**, 13226–13239 (1996).
- Bruchez, M. Jr, Moronne, M., Gin, P., Weiss, S. & Alivisatos, A. P. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
- Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = S, Se, Te) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
- Rossetti, R., Hull, R., Gibson, J. M. & Brus, L. E. Excited electronic states and optical spectra of ZnS and CdS crystallites in the ~15 to 50 Å size range: Evolution from molecular to bulk semiconducting properties. *J. Chem. Phys.* **82**, 552–559 (1985).
- Jenkins, R. & Snyder, R. L. *Introduction to X-ray Powder Diffractometry* (Wiley and Sons, New York, 1996).
- Wang, Y. R. & Duke, C. B. Atomic and electronic structure of ZnS cleavage surfaces. *Phys. Rev. B* **36**, 2736–2769 (1987).
- Raghavachari, K., Jakob, P. & Chabal, Y. J. Step relaxation and surface stress at H-terminated vicinal Si(111). *Chem. Phys. Lett.* **206**, 156–160 (1993).
- Rühl, E. *et al.* K-shell spectroscopy of Ar gas clusters. *J. Chem. Phys.* **98**, 6820–6826 (1993).
- Farges, J., de Feraudy, M. F., Raoult, B. & Torchet, G. Noncrystalline structure of argon clusters. II. Multilayer icosahedral structure of Ar_N clusters 50 < N < 750. *J. Phys. Chem.* **84**, 3491–3501 (1986).
- Muiliu, J. & Pakkanen, T. A. *Ab initio* study of small zinc sulfide crystallites. *Surf. Sci.* **364**, 439–452 (1996).
- Eichkorn, K. & Ahlrichs, R. Cadmium selenide semiconductor nanocrystals: A theoretical study. *Chem. Phys. Lett.* **288**, 235–242 (1998).
- Franceschetti, A., Fu, H., Wang, L. W. & Zunger, A. Many-body pseudopotential theory of excitons in InP and CdSe quantum dots. *Phys. Rev. B* **60**, 1819–1829 (1999).
- Leung, K. & Whaley, K. B. Surface relaxation in CdSe nanocrystals. *J. Chem. Phys.* **110**, 11012–11022 (1999).
- Rabani, E. Structure and electrostatic properties of passivated CdSe nanocrystals. *J. Chem. Phys.* **115**, 1493–1497 (2001).

22. McGinley, C. *et al.* Evidence for surface reconstruction on InAs nanocrystals. *Phys. Rev. B* **65**, 245308 (2002).
23. Meulenbergh, R. W. & Strouse, G. F. Vibrational analysis of nanocrystal CdSe surfaces: Evidence for surface reconstruction. *Abstr. Pap. Am. Chem. Soc.* **220**, 404-Phys (2000).
24. Hertl, W. Surface chemical properties of zinc sulfide. *Langmuir* **4**, 594–598 (1988).
25. Rockenburger, J. *et al.* The contributions of particle core and surface to strain, disorder and vibrations in thiolcapped CdTe nanocrystals. *J. Chem. Phys.* **108**, 7807–7815 (1998).
26. Head-Gordon, T. & Hura, G. Water structure from scattering experiments and simulation. *Chem. Rev.* **102**, 2651–2670 (2002).
27. Guinier, A. *X-Ray Diffraction in Crystals, Imperfect Crystals and Amorphous Bodies* (Freeman & Co., San Francisco, 1963).
28. Wright, K. & Jackson, R. A. Computer simulations of the structure and defect properties of zinc sulfide. *J. Mater. Chem.* **5**, 2037–2040 (1995).
29. Desgreniers, S., Beaulieu, L. & Lepage, I. Pressure-induced structural changes in ZnS. *Phys. Rev. B* **61**, 8726–8733 (2000).
30. de Leeuw, N. H. & Parker, S. C. Molecular-dynamics simulation of MgO surfaces in liquid water using a shell-model potential for water. *Phys. Rev. B* **58**, 13901–13908 (1998).
31. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
32. Stillinger, F. H. & Rahman, A. Revised central force potentials for water. *J. Chem. Phys.* **68**, 666–670 (1978).
33. Harris, D. J., Brodholt, J. P., Harding, J. H. & Sherman, D. M. Molecular dynamics simulation of aqueous ZnCl₂ solutions. *Mol. Phys.* **99**, 825–833 (2001).
34. Stevens, J. E., Chaudhuri, R. K. & Freed, K. F. Global three-dimensional potential energy surfaces of H₂S from the *ab initio* effective valence shell Hamiltonian method. *J. Chem. Phys.* **105**, 8754–8768 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Alivisatos for access to equipment, and W. Smith, T. R. Forester and D. Fincham for providing MD codes. EXAFS data were acquired on the DCM beamline at the UW-Madison Synchrotron Radiation Center (SRC), and we thank A. Jürgensen. WAXS data were acquired on beamline 11-ID-C at the Advanced Photon Source (APS), and we thank Y. Ren and M. Beno. HRTEM was performed at the National Center for Electron Microscopy, Berkeley, California. We also thank J. Rustad and G. Waychunas for discussions. This work was supported by the US Department of Energy (DOE), the Lawrence Berkeley National Laboratory LDRD, and the US National Science Foundation (NSF). The SRC is supported by the Division of Materials Research of the US NSF. Use of the Advanced Photon Source is supported by the US DOE, Office of Science, Office of Basic Energy Sciences.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.F.B. (jill@eps.berkeley.edu).

Controlling molecular deposition and layer structure with supramolecular surface assemblies

James A. Theobald¹, Neil S. Oxtoby², Michael A. Phillips¹,
Neil R. Champness² & Peter H. Beton¹

¹School of Physics and Astronomy, ²School of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK

Selective non-covalent interactions have been widely exploited in solution-based chemistry to direct the assembly of molecules into nanometre-sized functional structures such as capsules, switches and prototype machines^{1–5}. More recently, the concepts of supramolecular organization have also been applied to two-dimensional assemblies on surfaces^{6,7} stabilized by hydrogen bonding^{8–14}, dipolar coupling^{15–17} or metal co-ordination¹⁸. Structures realized to date include isolated rows^{8,13–15}, clusters^{9,10,18} and extended networks^{10–12,17}, as well as more complex multi-component arrangements¹⁶. Another approach to controlling surface structures uses adsorbed molecular monolayers to create preferential binding sites that accommodate individual target molecules^{19,20}. Here we combine these approaches, by using hydrogen bonding to guide the assembly of two types of molecules into a two-dimensional open honeycomb network that

then controls and templates new surface phases formed by subsequently deposited fullerene molecules. We find that the open network acts as a two-dimensional array of large pores of sufficient capacity to accommodate several large guest molecules, with the network itself also serving as a template for the formation of a fullerene layer.

We investigate an open honeycomb network formed when perylene tetra-carboxylic di-imide (PTCDI; Fig. 1a) is co-adsorbed with melamine (1,3,5-triazine-2,4,6-triamine; Fig. 1b) on a silver-terminated silicon surface. Melamine, which has a three-fold symmetry, forms the vertices of the network while the straight edges correspond to PTCDI. The network is stabilized by melamine–PTCDI hydrogen bonding. Melamine and PTCDI were chosen for this application because they are expected to exhibit much stronger hetero- as opposed to homo-molecular hydrogen bonding. As shown in Fig. 1, the compatibility of molecular geometries results in three hydrogen bonds per melamine–PTCDI pair, as compared

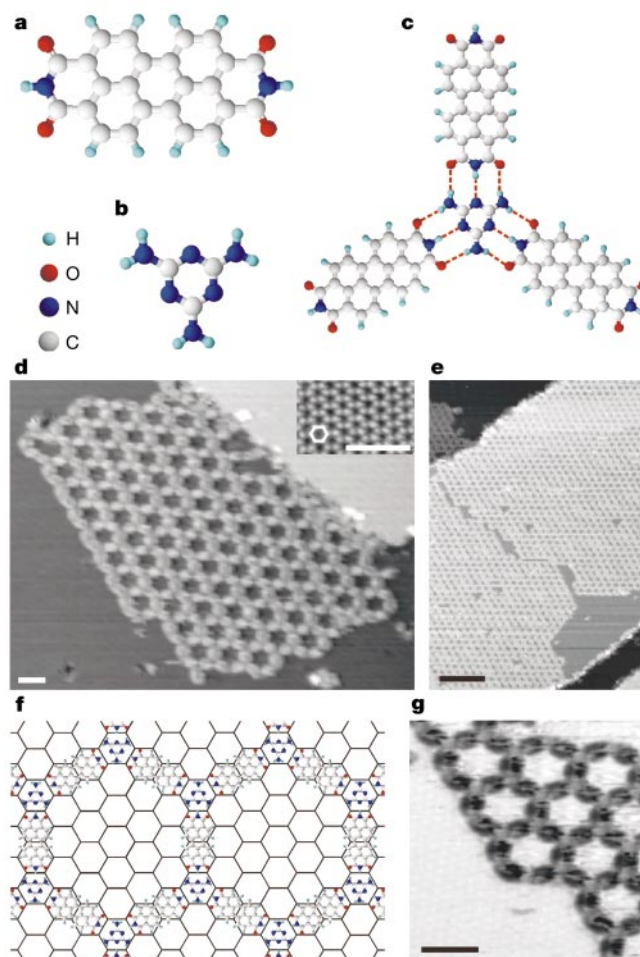


Figure 1 Self-assembly of a PTCDI-melamine supramolecular network. **a**, **b**, Chemical structure of PTCDI (**a**) and melamine (**b**). **c**, Schematic diagram of a PTCDI-melamine junction. Dotted lines represent the stabilizing hydrogen bonds between the molecules. **d**, STM image of a PTCDI-melamine network (sample voltage -2 V, tunnel current 0.1 nA). Inset, high-resolution view of the Ag/Si(111)- $\sqrt{3} \times \sqrt{3}R30^\circ$ substrate surface; the vertices and centres of hexagons correspond, respectively, to the bright (Ag trimers) and dark (Si trimers) topographic features in the STM image (surface lattice constant, $a_0 = 6.65$ Å; ref. 22). Scale bars, 3 nm. **e**, STM image of large-area network, with domains extending across terraces on the Ag/Si(111)- $\sqrt{3} \times \sqrt{3}R30^\circ$ surface (-2 V, 0.1 nA). Scale bar, 20 nm. **f**, Schematic diagram showing the registry of the network with the surface. **g**, Inverted contrast image (-2 V, 0.1 nA) of the network. Scale bar, 3 nm.

with just two for a PTCDI or melamine pair. The supramolecular synthon used in the melamine–PTCDI interaction has been shown to be very robust in solution-based supramolecular chemistry^{21,22}.

The network is prepared under ultra-high vacuum (UHV) conditions (base pressure $\sim 5 \times 10^{-11}$ torr). PTCDI and melamine were placed in effusion cells and sublimed (by heating to $\sim 360^\circ\text{C}$ and $\sim 100^\circ\text{C}$, respectively) onto a Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30° surface that had been prepared using standard procedures²³. The choice of substrate was motivated by previous studies which showed that molecules such as fullerenes, phthalocyanines and naphthalene tetracarboxylic di-imide (NTCDI, closely related to PTCDI)^{13,23–26} diffuse freely on the Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30° surface, and form islands in which the order is predominantly governed by inter-molecular interactions. Images of the surface were acquired using a scanning tunnelling microscope (STM) housed within the UHV system and operating at room temperature. The Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30° surface (Fig. 1d inset) has been studied extensively, and is described by the honeycomb-chain-trimer model in which each surface Si atom is bonded to one Ag atom^{27,28}. To simplify the subsequent discussion, we represent this

surface by a hexagonal network²⁵.

The first step in the formation of the network is deposition of 0.1–0.3 monolayers (ML) of PTCDI onto a surface held at room-temperature, after which close packed islands and short chains similar to those reported in previous studies of PTCDI and NTCDI^{13,29} are observed. Melamine is then deposited while the sample is annealed at $\sim 100^\circ\text{C}$. STM images of the resulting molecular network are shown in Fig. 1. The network has principal axes at 30° to those of the Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30° surface, and a lattice constant $3\sqrt{3}a_0 = 34.6 \text{ \AA}$. In images with inverted topographic contrast (Fig. 1g), the positions of the two molecules may be clearly discerned, with melamine and PTCDI forming respectively the vertices and edges of the network, which is stabilized by the hydrogen bonding illustrated in Fig. 1c. The maximum surface coverage of the network observed to date is 50% (sample shown in large-area STM image, Fig. 1e). The registry of the molecules with respect to the underlying surface has been determined, and is shown schematically in Fig. 1f. Numerical calculations (Austin method 1³⁰) give the centre-to-centre spacing of a single PTCDI–melamine pair as $10.2 \text{ \AA}$, close to the observed separation of $1.5a_0 = 9.98 \text{ \AA}$. Thus the calculated melamine–melamine separation has a near-commensurability with the surface lattice.

Annealing provides sufficient thermal energy for molecules to detach from PTCDI islands and diffuse across the surface. These PTCDI molecules interact with melamine to nucleate the hexagonal

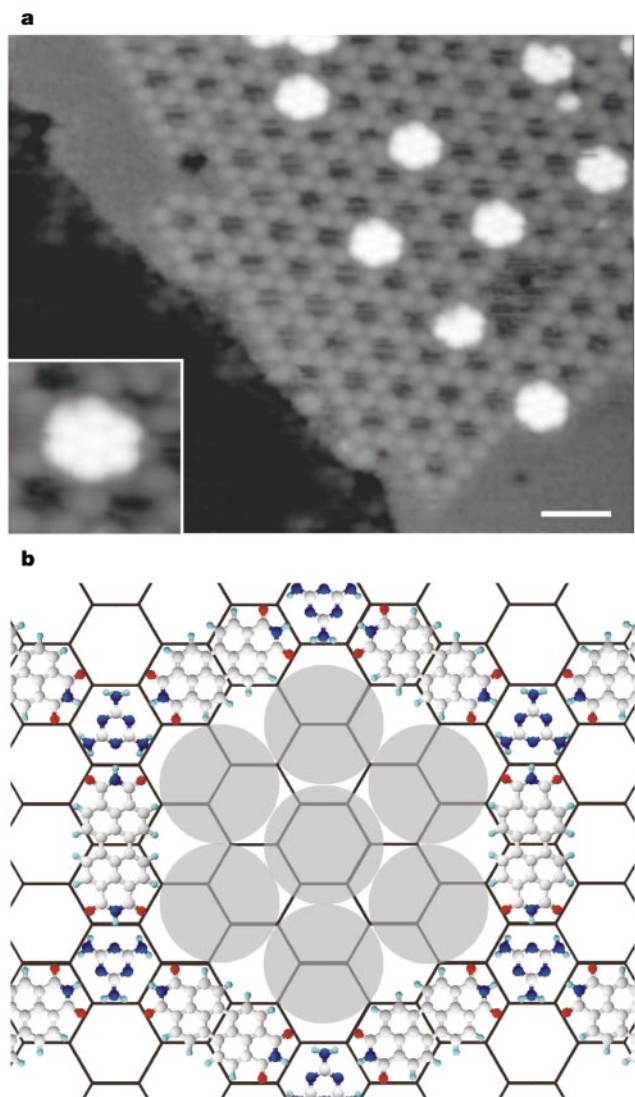


Figure 2 Images of C₆₀ heptamers trapped within the ‘nanoscale vessels’. **a**, STM image (-2 V , 0.1 nA) of C₆₀ heptamers on a PTCDI–melamine network. Inset, high-resolution view showing an individual cluster. Scale bar, 5 nm . **b**, Schematic diagram of a C₆₀ heptamer.

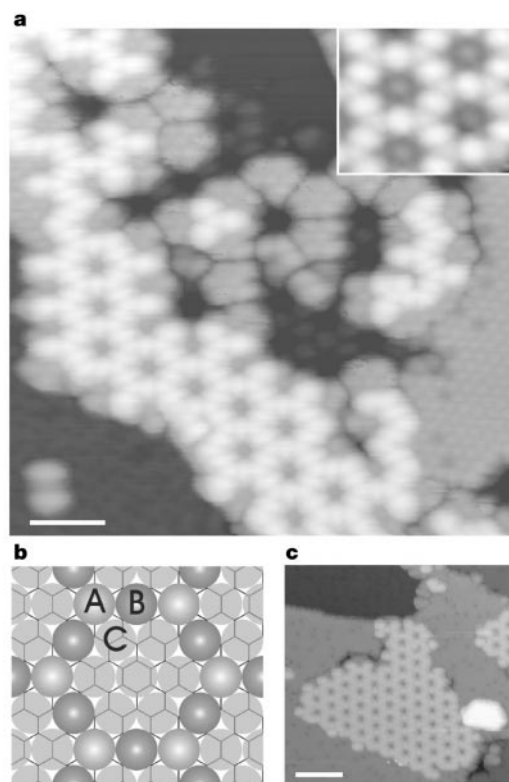


Figure 3 Images of the C₆₀ honeycomb network. **a**, Image showing the PTCDI–melamine lattice, C₆₀ heptamers and the raised C₆₀ honeycomb network (-2 V , 0.1 nA). Scale bar, 5 nm . Inset, high-resolution view showing the C₆₀ honeycomb network. **b**, Schematic diagram showing the registry of the raised C₆₀ network with the surface. The hexagonal network of C₆₀ molecules, marked A and B, sit directly above the melamine and PTCDI, respectively. These molecules are raised with respect to the heptamers by 2.1 and $2.8 \text{ \AA}$, respectively. The elevation of molecules A and B results in an increase in their separation from molecules at the heptamer edge (marked C) to $\sim 10.3 \text{ \AA}$ (assuming no lateral relaxation of the heptamer molecules — this cannot be resolved from STM images). **c**, Low defect termination of a PTCDI–melamine lattice with C₆₀ (-2 V , 0.1 nA). Scale bar, 10 nm .

network, which then grows through further capture of diffusing molecules. If deposition of PTCDI is continued or restarted during formation of the network, the pores become 'filled' with other PTCDI molecules, presumably by their direct impinging on the pore area. The open network may only be formed from PTCDI molecules released from pre-formed islands and thus constrained to two-dimensional diffusion on the surface. The stepwise introduction of the two components is thus vital for the synthesis of the PTCDI-melamine network.

Hexagonal networks assembled from a single molecular species have been reported previously^{10–12,17}. The open areas in such networks are small, comparable in size to that of a single molecular component of the supramolecular array¹¹. In contrast, the incorporation of linear PTCDI edge molecules into a bimolecular assembly yields pores that are much larger than the constituent building blocks of the network and thus capable of serving as traps, or vessels, for the co-location of several large molecules. We demonstrate this potential by subliming C₆₀ onto the hexagonal network. As shown in Fig. 2a, which shows an STM image acquired after deposition of 0.03 ML of C₆₀, heptameric C₆₀ clusters with a compact hexagonal arrangement of the individual molecules form within the pores. Clusters formed in different pores are aligned, and are all oriented parallel to the principal axes of the Si(111) surface. The molecular arrangement of the heptamers has been deduced from STM images (Fig. 2b). Clusters of fewer molecules are also observed. For example, there are clusters of six molecules in Fig. 2a and clusters of 2–5 molecules have also been observed, while many pores remain empty for this coverage of C₆₀.

Previous studies of C₆₀ on Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30° observed extended monolayers and multilayer islands, but not isolated hexagonal heptameric clusters such as those in Fig. 2^{25,26}. Moreover, whereas the most stable C₆₀ islands on Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30° were previously found to be hexagonally close-packed, their principal axes were misoriented with respect to the underlying surface^{25,26} and the clusters were thus off-axis, unlike the on-axis heptamers observed here. These observations clearly indicate that the heptamers are stabilized by the PTCDI-melamine network.

The fraction of pores containing adsorbed molecules and stabilized heptameric clusters increases with increasing C₆₀ coverage. In addition, we observe adsorption of C₆₀ directly above the PTCDI and melamine molecules, thus reproducing the underlying hexagonal network. Figure 3a shows STM images of second-layer C₆₀, heptamers and the PTCDI-melamine network in close proximity. A further increase in C₆₀ coverage results in a near-perfect termination of the second layer (see STM images in Fig. 3a, c). An array of C₆₀ molecules sits directly above the melamine-PTCDI network (Fig. 3b), with the lateral positions within this array corresponding exactly to those of a hexagonally close-packed layer. The elevation of the hexagonal network's constituent molecules increases their separation from molecules at the heptamer edge, and this arrangement thus constitutes a new fullerene surface phase that is controlled and templated by the underlying hydrogen-bonded network.

Further deposition of C₆₀, up to a total of 3 ML, does not lead to the formation of additional fullerene layers on the hexagonal network. We attribute this observation to the termination layer (Fig. 3) lacking sites for the stable nucleation of additional layers. In particular, we find that the largest close-packed cluster that could form in a subsequent layer would contain only three molecules, rendering the cluster kinetically unstable at room temperature. Molecules incident on the hexagonal network diffuse to regions of the Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30° surface that are not yet terminated, and where they are incorporated in close-packed fullerene islands that act as a sink for incident C₆₀. This suppression of nucleation, which is likely to depend on preparation conditions such as substrate temperature, accounts for the high degree of perfection of the termination observed in Fig. 3c.

Our results show that large pore areas can induce the co-location

of several molecules in a 'nanoscale vessel'. Such pores might also prove useful in achieving the controlled co-location of a wider range of guest species, which could in turn lead to the promotion of local chemical interactions, controlled polymerization and the formation of complex supramolecular surface structures. □

Received 13 March; accepted 17 July 2003; doi:10.1038/nature01915.

- Lehn, J. M. Toward complex matter: Supramolecular chemistry and self-organization. *Proc. Natl Acad. Sci. USA* **99**, 4763–4768 (2002).
- Balzani, V., Credi, A., Raymo, F. M. & Stoddart, J. F. Artificial molecular machines. *Angew. Chem. Int. Edn Engl.* **39**, 3349–3391 (2000).
- Reinhoudt, D. N. & Crego-Calama, M. Synthesis beyond the molecule. *Science* **295**, 2403–2407 (2002).
- Fujita, M., Fujita, N., Ogura, K. & Yamaguchi, K. Spontaneous assembly of ten components into two interlocked, identical coordination cages. *Nature* **400**, 52–55 (1999).
- Seeman, N. C. DNA in a material world. *Nature* **421**, 427–431 (2003).
- Hecht, S. Welding, organizing, and planting organic molecules on substrate surfaces — Promising approaches towards nanoarchitectonics from the bottom up. *Angew. Chem. Int. Edn Engl.* **42**, 24–26 (2003).
- De Feyter, S. & De Schryver, F. C. Two-dimensional supramolecular self-assembly probed by scanning tunnelling microscopy. *Chem. Soc. Rev.* **32**, 139–150 (2003).
- Barth, J. V. et al. Building supramolecular nanostructures at surfaces by hydrogen bonding. *Angew. Chem. Int. Edn Engl.* **39**, 1230–1234 (2000).
- Bohringer, M., Morgenstern, K., Schneider, W. D. & Berndt, R. Separation of a racemic mixture of two-dimensional molecular clusters by scanning tunneling microscopy. *Angew. Chem. Int. Edn Engl.* **38**, 821–823 (1999).
- Furukawa, M., Tanaka, H. & Kawai, T. Formation mechanism of low-dimensional superstructure of adenine molecules and its control by chemical modification: A low-temperature scanning tunneling microscopy study. *Surf. Sci.* **445**, 1–10 (2000).
- Griessl, S., Lackinger, M., Edelwirth, M., Hietschold, M. & Heckl, W. M. Self-assembled two-dimensional molecular host-guest architectures from trimesic acid. *Single Mol.* **3**, 25–31 (2002).
- Dmitriev, A., Lin, N., Weckesser, J., Barth, J. V. & Kern, K. Supramolecular assemblies of trimesic acid on a Cu(100) surface. *J. Phys. Chem. B* **106**, 6907–6912 (2002).
- Keeling, D. L. et al. Assembly and processing of hydrogen bond induced supramolecular nanostructures. *Nano Lett.* **3**, 9–12 (2003).
- Chen, Q., Frankel, D. J. & Richardson, N. V. Self-assembly of adenine on Cu(110) surfaces. *Langmuir* **18**, 3219–3225 (2002).
- Yokoyama, T., Yokoyama, S., Kamikado, T., Okuno, Y. & Mashiko, S. Selective assembly on a surface of supramolecular aggregates with controlled size and shape. *Nature* **413**, 619–621 (2001).
- de Wild, M. et al. A novel route to molecular self-assembly: Self-intermixed monolayer phases. *ChemPhysChem* **3**, 881–885 (2002).
- Berner, S. et al. Time evolution analysis of a 2D solid-gas equilibrium: A model system for molecular adsorption and diffusion. *Chem. Phys. Lett.* **348**, 175–181 (2001).
- Lin, N., Dmitriev, A., Weckesser, J., Barth, J. V. & Kern, K. Real-time single-molecule imaging of the formation and dynamics of coordination compounds. *Angew. Chem. Int. Edn Engl.* **41**, 4779–4783 (2002).
- Gimzewski, J. K., Jung, T. A., Cuberes, M. T. & Schlittler, R. R. Scanning tunneling microscopy of individual molecules: Beyond imaging. *Surf. Sci.* **386**, 101–114 (1997).
- Cuberes, M. T., Schlittler, R. R. & Gimzewski, J. K. Room temperature supramolecular repositioning at molecular interfaces using a scanning tunneling microscope. *Surf. Sci.* **371**, L231–L234 (1997).
- Paraschiv, V. et al. Molecular "chaperones" guide the spontaneous formation of a 15-component hydrogen-bonded assembly. *J. Am. Chem. Soc.* **124**, 7638–7639 (2002).
- Prins, L. J., Reinhoudt, D. N. & Timmerman, P. Noncovalent synthesis using hydrogen bonding. *Angew. Chem. Int. Edn Engl.* **40**, 2383–2426 (2001).
- Wan, K. J., Lin, X. F. & Nogami, J. Surface reconstructions in the Ag/Si(111) system. *Phys. Rev. B* **47**, 13700–13712 (1993).
- Upward, M. D., Beton, P. H. & Moriarty, P. Adsorption of cobalt phthalocyanine on Ag terminated Si(111). *Surf. Sci.* **441**, 21–25 (1999).
- Upward, M. D., Moriarty, P. & Beton, P. H. Double domain ordering and selective removal of C₆₀ on Ag/Si(111)- $\sqrt{3} \times \sqrt{3}$ R30°. *Phys. Rev. B* **56**, R1704–R1707 (1997).
- Nakayama, T., Onoe, J., Takeuchi, K. & Aono, M. Weakly bound and strained C₆₀ monolayer on the Si(111)- $\sqrt{3} \times \sqrt{3}$ R30°-Ag substrate surface. *Phys. Rev. B* **59**, 12627–12631 (1999).
- Takahashi, T., Nakatani, S., Okamoto, N., Ishikawa, T. & Kikuta, S. A Study of the Si(111)- $\sqrt{3} \times \sqrt{3}$ -Ag surface by transmission-X-ray diffraction and X-ray-diffraction topography. *Surf. Sci.* **242**, 54–58 (1991).
- Katayama, M., Williams, R. S., Kato, M., Nomura, E. & Aono, M. Structure analysis of the Si(111)- $\sqrt{3} \times \sqrt{3}$ R30°-Ag Surface. *Phys. Rev. Lett.* **66**, 2762–2765 (1991).
- Uder, B., Ludwig, C., Petersen, J., Gompf, B. & Eisenmenger, W. STM characterization of organic molecules on H-terminated Si(111). *Z. Phys. B* **97**, 389–390 (1995).
- Dewar, M. J. S., Zebisch, E. G., Healy, E. F. & Stewart, J. J. P. The development and use of quantum-mechanical molecular models, 76. AM1 — a new general-purpose quantum-mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).

Acknowledgements We thank P. Moriarty and J. O'Shea for discussions, and N. Besley and P. Gill for advice on numerical calculations. This work was supported by the UK Engineering and Physical Sciences Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.R.C. (Neil.Champness@nottingham.ac.uk) or P.H.B. (Peter.Beton@nottingham.ac.uk).

Stability of the body-centred-cubic phase of iron in the Earth's inner core

Anatoly B. Belonoshko^{1,2}, Rajeev Ahuja³ & Börje Johansson^{1,3}

¹Applied Materials Physics, Department of Materials Science and Engineering, The Royal Institute of Technology, SE-100 44 Stockholm, Sweden

²Condensed Matter Theory Group, Stockholm Center for Physics, Astronomy and Biotechnology, Department of Physics, The Royal Institute of Technology, SE-106 91 Stockholm, Sweden

³Condensed Matter Theory Group, Department of Physics, Uppsala University, Box 530, SE-751 21, Uppsala, Sweden

Iron is thought to be the main constituent of the Earth's core¹, and considerable efforts^{2–14} have therefore been made to understand its properties at high pressure and temperature. While these efforts have expanded our knowledge of the iron phase diagram, there remain some significant inconsistencies, the most notable being the difference between the 'low' and 'high' melting curves¹⁵. Here we report the results of molecular dynamics simulations of iron based on embedded atom models fitted to the results of two implementations of density functional theory. We tested two model approximations and found that both point to the stability of the body-centred-cubic (b.c.c.) iron phase at high temperature and pressure. Our calculated melting curve is in agreement with the 'high' melting curve, but our calculated phase boundary between the hexagonal close packed (h.c.p.) and b.c.c. iron phases is in good agreement with the 'low' melting curve. We suggest that the h.c.p.–b.c.c. transition was previously misinterpreted as a melting transition, similar to the case of xenon^{16–18}, and that the b.c.c. phase of iron is the stable phase in the Earth's inner core.

Remarkably, the iron melting temperature at the highest static pressure ($3,850 \pm 200$ K at 197 GPa) reached⁷ is, within experimental error, essentially the same as the pressure–temperature (P – T)

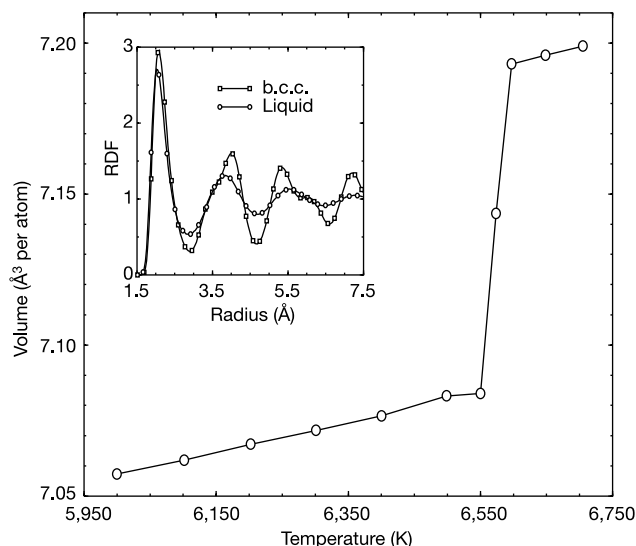


Figure 1 Temperature dependence of the volume of iron. The volume was calculated using the two-phase¹¹ molecular dynamics method with the EAM PAW²¹ model for the interactions. The starting configuration was b.c.c. + liquid. At $T \leq 6,550$ K the liquid part solidified into the b.c.c. structure. At $T \geq 6,600$ K the b.c.c. part melted. The insertion shows the radial distribution function (RDF) at a temperature of 6,550 K (b.c.c.) and 6,600 K (liquid).

conditions of the discontinuity observed in shock-wave experiments³ ($4,350 \pm 300$ K at 200 GPa). In the shock-wave experiment this discontinuity was interpreted as due to a solid–solid transition, which was afterwards attributed² to a h.c.p.–b.c.c. transition. Later, the shock-wave experiment was repeated¹⁹ and it was claimed that the original finding³ was not confirmed. However, this discrepancy is probably exaggerated²⁰. The difference can be explained by possible hysteresis and the differing compositions of the samples²⁰.

The possibility of a high- P – T b.c.c. iron phase was rejected on the basis of two arguments coming from *ab initio* calculations^{4,5} at $T = 0$ K, namely: (1) the large energy difference between the h.c.p. and b.c.c. phases and (2) the calculated mechanical instability of the b.c.c. structure. Recently, using the same experimental technique as in the case of Fe⁷, a peculiar transition was observed¹⁶ in Xe. This transition was interpreted as melting. However, it was later demonstrated that the transition is more probably an f.c.c.–b.c.c. transition¹⁸. If true, we have a case where the solid–solid transition is misinterpreted as melting. If this is possible in the case of Xe, it could also be the case for Fe. Therefore, the 'low' melting curve⁷ at high pressure might in fact represent a solid–solid transition.

To investigate this possibility, we carried out a molecular dynamics study based on two non-empirical embedded atom models (EAM) fitted to the results of first-principles FPLMTO¹¹ and PAW²¹ calculations of iron energies respectively. Both models correctly predict thermodynamic stability of the h.c.p. phase of iron relative to the b.c.c. phase at high P and $T = 0$ K. The EAM FPLMTO model reproduces the dynamical instability of b.c.c. phase at high P in agreement with previous calculations^{4,5}. The EAM PAW model²¹ was created as a fine-tuning of our EAM FPLMTO¹¹ model for the energies of the iron phases in a narrow

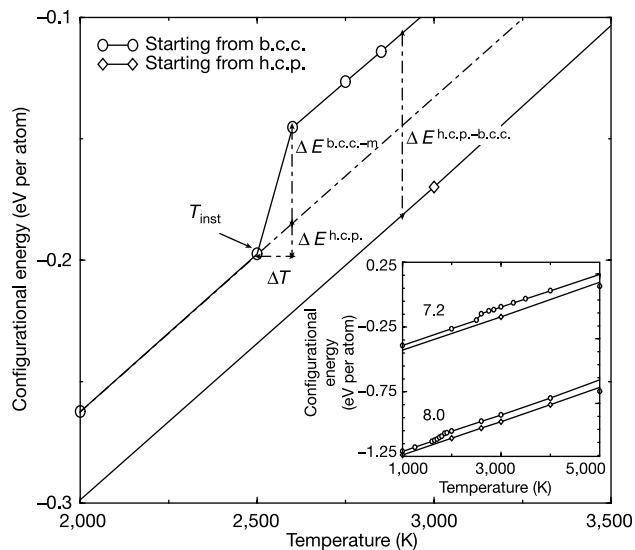


Figure 2 Energies of the iron phases at constant volume as a function of temperature. Energies are calculated using the molecular dynamics method with the EAM FPLMTO¹¹ model for the interaction. Simulation with h.c.p. as a starting configuration always ended in an h.c.p. structure. Simulations starting from the b.c.c. structure resulted in different structures depending on temperature. The designations illustrate the calculations of entropy (equation (1)). The insertion shows our calculated isochores for two volumes (7.2 and 8.0 Å³ per atom). Similar isochores were calculated also at volumes 6.6, 7.6, 8.5, and 9.0 Å³ per atom. The energies of the h.c.p. and b.c.c. iron phases are calculated using configurations of 64 and 54 thousand atoms, respectively. Test runs with a considerably higher number of atoms demonstrated that the value of T_{eq} (equation (3)) does not depend on the number of atoms in the simulation.

P and T range at about 320 GPa and 6,000 K.

Using the two-phase method, the melting temperature of the h.c.p. phase was calculated²¹ to be 6,300 K at $P = 323.5$ GPa. We also used the two-phase¹¹ method and the same (EAM PAW) model to calculate the melting temperature of the b.c.c. phase at the same (323.5 GPa) pressure. We found that b.c.c. iron melts at 6,600 K when the pressure is 323.5 GPa (Fig. 1). This is approximately 300 K higher than the calculated²¹ melting temperature for the h.c.p. phase. Hence, according to the EAM PAW model, the b.c.c. phase is more stable than the h.c.p. phase. At low T and high P the EAM PAW model gives rise to the dynamically stable b.c.c. iron phase, while from *ab initio* calculations^{4,5} it is found to be mechanically unstable at $T = 0$ K. Therefore, for the calculation of the iron b.c.c.–h.c.p. boundary, we decided to use our EAM FPLMTO model, which might be less precise than EAM PAW model at a particular pressure, but performs nicely over a broad P – T range¹¹.

We calculated the energies of the h.c.p. and b.c.c. iron phases as a function of T at six volumes (Fig. 2). We noticed that on cooling the b.c.c. phase transforms into a martensitic^{17,22} structure in a rather narrow temperature range. This structure consists of h.c.p. and f.c.c. fragments which have different orientations and are connected to each other by either low-energy twin boundaries or high-energy interfaces²². It has the same entropy as the h.c.p. structure (at high P only the number of ways in which energy can be distributed between the nearest neighbours is important), which is evident also from the equal slopes of the energy (E)– T dependence of the martensitic and h.c.p. structures (Fig. 2). The martensitic structure was observed in numerous computer and real experiments^{17,22}. Therefore, this is

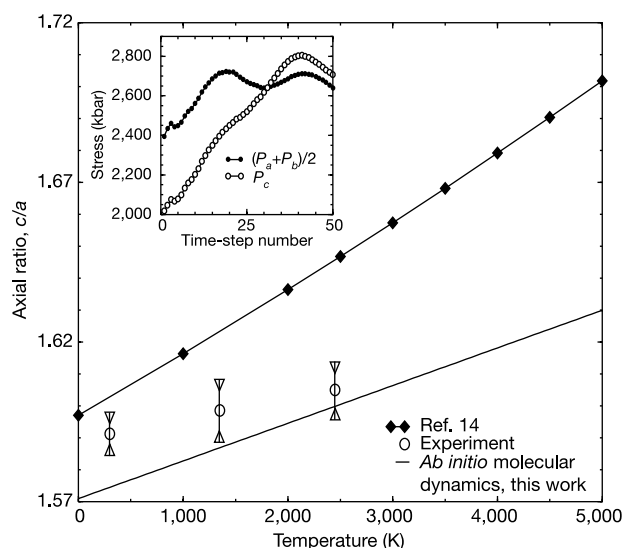


Figure 3 The axial ratio of h.c.p. iron as a function of temperature. The ratio is calculated at a constant volume ($7.2 \text{ \AA}^3$ per atom) using *ab initio* molecular dynamics¹² and compared with experimental^{15,26} (taken at pressures above 1.5 Mbar, triangles indicate error bars) and calculated previously¹⁴ (interpolated) data. The insertion demonstrates the time dependence of stresses in the basal plane and along the c axis at $T = 5,000$ K, $V = 7.2 \text{ \AA}^3$ per atom and $c/a = 1.65$. It is clear that during the first 50 time-steps, while thermal equilibrium is not yet reached, P_c (stress along the c axis) becomes larger than P_a , providing the incorrect impression that the c/a ratio must be increased to obtain equal stresses. The neglect of thermal equilibrium led to a very high calculated c/a ratio¹⁴, where all atoms in a computational cell were frozen except one. In our method all atoms are allowed to move, which seems to lead to much better agreement with experiment, particularly when the temperature increases. Our data are approximated by a straight line for clarity. At 5,000 K the ratio is close to ideal ($c/a = 1.633$).

probably the phase which does not change its properties with time, and calculation of entropy of this phase is legitimate. We point out that the P – T conditions where the “weak change of laser adsorption”⁶ on temperature cycling was observed⁶ (and attributed to some yet unknown phase transition) is close to the P – T conditions of the b.c.c.–martensitic structure transition obtained in our present simulations. It is possible that the variety of new iron phases reported recently (double-h.c.p.⁷ and orthorhombic⁸ structures) might be explained by the appearance of the martensitic structure. The dependence of the energy on T is practically linear (Fig. 2) and is in compliance with the Dulong–Petit law. Identical heat capacities of martensitic and h.c.p. phases up to the T_{inst} (the temperature of the transition of the b.c.c. structure into the martensitic structure, Fig. 2) and of the b.c.c. and h.c.p. phases above the temperature of $T_{\text{inst}} + \Delta T$ (Fig. 2) allow us greatly to simplify the calculations of the entropy difference (ΔS) between the b.c.c. and h.c.p. phases. The kinetic energy obviously cancels at the same T , so we can write:

$$\Delta S = \int_{T_{\text{inst}}}^{T_{\text{inst}} + \Delta T} \left[\frac{\left(\frac{\Delta E^{\text{b.c.c.-m}} + \Delta E^{\text{h.c.p.}}}{\Delta T} \right) - \frac{\Delta E^{\text{h.c.p.}}}{\Delta T}}{T} \right] dt \quad (1)$$

where $\Delta E^{\text{b.c.c.-m}}$ is the potential energy difference between the

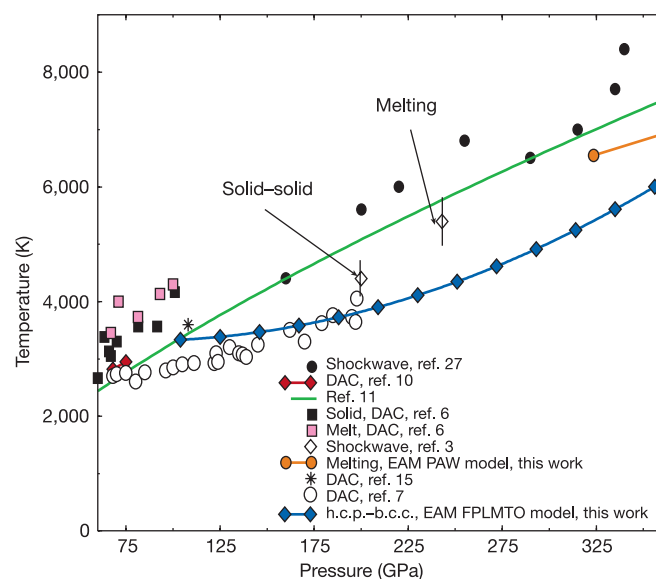


Figure 4 High-pressure iron phase diagram. Theoretical and experimental data on the stability of h.c.p., b.c.c. and liquid phases of iron at high pressure are shown by curves and unconnected symbols respectively. Calculated new data are shown for the h.c.p.–b.c.c. transition (blue diamonds, EAM FPLMTO¹¹ model) and melting of the b.c.c. phase (orange circles, EAM PAW²¹ model). The calculated h.c.p.–b.c.c. transition almost coincides with the “low” melting curve⁷ (open circles) and lies within the error bars (shown as vertical lines) for the solid–solid transition as measured in shock-wave experiments³. The melting curve, which we calculated before¹¹ (solid green line), is in agreement with the shock-wave melting temperatures^{3,27} (the diamond with vertical bar and the black circles) and with experimental data obtained by the diamond anvil cell (DAC) technique at moderate pressures (red diamonds¹⁰). The latest measurements with improved technique¹⁵ gave a melting temperature (star symbol) which agrees very well with what we predicted earlier¹¹ (green line). Of the two recently calculated melting curves^{13,21} (not shown), one is close to our calculated melting curve^{11,21} (orange circles) and the other¹³ is close to our presently calculated h.c.p.–b.c.c. transition (blue diamonds). The experimental evidence for the new solid–solid transition²⁸ (not shown) at $P = 125$ GPa and $T = 3,300$ K is exactly where we predict the h.c.p.–b.c.c. transition (blue diamonds).

b.c.c. and martensitic structure, and $\Delta E^{\text{h.c.p.}}$ is the increment of the potential energy of the h.c.p. phase when T changes by ΔT . Because ΔT is small relative to T_{inst} and $\ln(1+x) \approx x$ at small x , equation (1) can be simplified, and $\Delta S^{\text{h.c.p.-b.c.c.}} = \Delta E^{\text{b.c.c.-m}}/T_{\text{inst}}$. From the equilibrium condition at constant volume (F is the Helmholtz free energy):

$$\Delta F^{\text{h.c.p.-b.c.c.}} = \Delta E^{\text{h.c.p.-b.c.c.}} - T\Delta S^{\text{h.c.p.-b.c.c.}} = 0 \quad (2)$$

we obtain the equilibrium temperature (T_{eq}) between the b.c.c. and h.c.p. phases:

$$T_{\text{eq}} = T_{\text{inst}} \Delta E^{\text{h.c.p.-b.c.c.}} / \Delta E^{\text{b.c.c.-m}} \quad (3)$$

Interestingly, direct calculation of the entropy from the autocorrelation function gives very high temperatures for the h.c.p.-b.c.c. transition in zirconium compared to the experimental value (about 1,800 K as compared to 1,135 K), whereas a calculation for the martensitic structure has allowed very reasonable agreement (1,350 K as compared to 1,135 K) using the EAM model²².

We emphasize that the energy difference between the h.c.p. and b.c.c. structures at $T = 0$ K has little bearing on the stability of the b.c.c. phase at high T , because the b.c.c. phase does not exist at low T owing to the dynamic instability, and, therefore, integration of the entropy from 0 K to high T is impossible. However, it is important to verify the ability of our model to provide correct $\Delta E^{\text{h.c.p.-b.c.c.}}$ at high T . Applying an *ab initio* molecular dynamics method, as described before¹², we calculated $\Delta E^{\text{h.c.p.-b.c.c.}}$ at $T = 5,000$ K and volume $V = 7.2 \text{ \AA}^3$ per atom. We obtained $\Delta E^{\text{h.c.p.-b.c.c.}} = 0.102 \pm 0.041$ eV per atom, which agrees very well with our model (Fig. 2). In our calculations at high T we used the axial ratio $c/a = 1.63$, because we found (Fig. 3) that this value is what we obtain both from the EAM model and *ab initio* molecular dynamics simulations.

Having obtained the temperature and volumes for the equilibrium points, we calculated pressures of the h.c.p.-b.c.c. transition. These P - T points were then fitted with a polynomial, as shown in Fig. 4 along with the calculated EAM PAW melting points. The new data are compared with experimental data and our previously calculated iron melting curve. The calculation for the b.c.c. phase does not change the melting temperatures significantly, so our previously calculated melting curve is sufficient for the general comparison. Figure 4 shows that our EAM FPLMTO model allows us to reproduce all sets of data within experimental error. The agreement between the “low” melting curve and the calculated h.c.p.-b.c.c. phase boundary is extremely good, similar to the agreement between the “melting” curve¹⁶ and the f.c.c.-b.c.c. boundary in Xe¹⁸ at high P and T .

We obtain the triple h.c.p.-b.c.c.-liquid point at almost the same pressure (105 GPa), as was suggested in ref. 7 for the h.c.p.-f.c.c.-liquid point. More recent measurements¹⁵ now place this triple point at roughly 60 GPa. Even such a delicate feature of the experimental curve (open circles, Fig. 4) as downward inflection is reproduced. We also note that it is quite unusual for a melting curve to be inflected downward.

Thus, with such good agreement between the calculated h.c.p.-b.c.c. phase boundary (Fig. 4) and the experimental “low” melting curve⁷, we suggest that the experimental ‘low’ melting curve might, in fact, represent the h.c.p.-b.c.c. transition. The calculated h.c.p.-b.c.c. transition is also in agreement (Fig. 4) with the solid-solid transition evidenced by shock-wave experiments³. The P - T conditions in the Earth’s inner core are clearly above the calculated h.c.p.-b.c.c. boundary; therefore, if there is not another, even more stable solid phase, iron in the Earth’s inner core must be stable in the b.c.c. phase.

The model, based on first-principles calculations without any reference to experimental data, thus resolves the conflict between

two contradictory experimental sets of data. We note that our EAM FPLMTO model does not account for the electronic entropy. However, if electronic entropy is of any importance for the relative stability of h.c.p. and b.c.c., it would stabilize the b.c.c. phase even at lower temperature, as it is larger in the b.c.c.²³ than in the h.c.p. phase. We recommend detailed calculations of the b.c.c. iron phase elastic anisotropy, to discriminate between alternative models^{14,24} of the seismic anisotropy of the Earth’s inner core. In addition, consideration of the b.c.c. phase, which is less dense than the h.c.p. phase, might substantially diminish the amount of light elements in the Earth’s inner core to match its density²⁵. □

Received 11 October 2002; accepted 28 July 2003; doi:10.1038/nature01954.

1. Birch, F. Elasticity and constitution of the Earth’s interior. *J. Geophys. Res.* **57**, 227–286 (1952).
2. Ross, M., Young, D. A. & Grover, R. Theory of the iron phase diagram at earth core conditions. *J. Geophys. Res.* **95**, 21713–21716 (1990).
3. Brown, J. M. & McQueen, R. G. Grüneisen parameter, and elasticity for shocked iron between 77 GPa and 400 GPa. *J. Geophys. Res.* **91**, 7485–7494 (1986).
4. Stixrude, L., Cohen, R. E. & Singh, D. J. Iron at high pressure: Linearized-augmented-plane-wave computations in the generalized-gradient approximation. *Phys. Rev. B* **50**, 6442–6445 (1994).
5. Söderlind, P., Moriarty, J. A. & Wills, J. M. First-principles theory of iron up to earth-core pressures: Structural, vibrational, and elastic properties. *Phys. Rev. B* **53**, 14063–14072 (1996).
6. Williams, Q., Knittle, E. & Jeanloz, R. The high pressure melting curve of iron: A technical discussion. *J. Geophys. Res.* **96**, 2171–2184 (1991).
7. Boehler, R. Temperatures in the Earth’s core from melting-point measurements of iron at high-static pressures. *Nature* **363**, 534–536 (1993).
8. Saxena, S. K. *et al.* Synchrotron x-ray study of iron at high pressure and temperature. *Science* **269**, 1703–1704 (1995).
9. Andrault, D., Fiquet, G., Kunz, M., Vissekas, F. & Hausermann, D. The orthorhombic structure of iron: An *in situ* study at high temperature and high pressure. *Science* **278**, 831–834 (1997).
10. Shen, G., Mao, H. K., Hemley, R. J., Duffy, T. S. & Rivers, M. L. Melting and crystal structure of iron at high pressures and temperatures. *Geophys. Res. Lett.* **25**, 373–376 (1998).
11. Belonoshko, A. B., Ahuja, R. & Johansson, B. Quasi-*ab initio* molecular dynamic study of Fe melting. *Phys. Rev. Lett.* **84**, 3638–3641 (2000).
12. Alfé, D., Gillan, M. J. & Price, G. D. The melting curve of iron at the pressures of the Earth’s core from *ab initio* calculations. *Nature* **401**, 462–464 (1999).
13. Laio, A., Bernard, S., Chiarotti, G. L., Scandolo, S. & Tosatti, E. Physics of iron at Earth’s core conditions. *Science* **287**, 1027–1030 (2000).
14. Steinle-Neumann, G., Stixrude, L., Cohen, R. E. & Gülseren, O. Elasticity of iron at the temperature of the Earth’s inner core. *Nature* **413**, 57–60 (2001).
15. Hemley, R. J. & Mao, H.-K. In situ studies of iron under pressure: New windows on the earth’s core. *Int. Geol. Rev.* **43**, 1–30 (2001).
16. Boehler, R., Ross, M., Söderlind, P. & Boercker, D. B. High-pressure melting curves of argon, krypton, and xenon: Deviation from corresponding states theory. *Phys. Rev. Lett.* **86**, 5731–5734 (2001).
17. Errandonea, D., Schwager, B., Boehler, R. & Ross, M. Phase behavior of krypton and xenon to 50 GPa. *Phys. Rev. B* **65**, 214110 (2002).
18. Belonoshko, A. B., Ahuja, R. & Johansson, B. Molecular dynamics study of melting and f.c.c.-b.c.c. transitions in Xe. *Phys. Rev. Lett.* **87**, 165505 (2001).
19. Nguyen, J. H. & Holmes, N. C. Iron sound velocities in shock wave experiments up to 400 GPa. *Eos* **79**, F846 (1998).
20. Boness, D. A. in *Shock Compression of Condensed Matter-1999* (eds Furnish, M. D., Chhabildas, L. C. & Hixson, R. S.) 77–80 (American Institute of Physics, 2000).
21. Alfé, D., Gillan, M. J. & Price, G. D. Complementary approaches to the *ab initio* calculation of melting properties. *J. Chem. Phys.* **116**, 6170–6177 (2002).
22. Pinsook, U. Molecular dynamics study of vibrational entropy in b.c.c. and h.c.p. zirconium. *Phys. Rev. B* **66**, 024109 (2002).
23. Matsui, M. & Anderson, O. L. The case for a body-centered cubic phase (α') for iron at inner core conditions. *Phys. Earth Planet. Inter.* **103**, 55–62 (1997).
24. Singh, S. C., Taylor, M. A. & Montagner, J. P. On the presence of liquid in Earth’s inner core. *Science* **287**, 2471–2474 (2000).
25. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
26. Dubrovinsky, L. S. *et al.* In situ x-ray study of thermal expansion and phase transition of iron at multimegabar pressure. *Phys. Rev. Lett.* **84**, 1720–1723 (2000).
27. Yoo, C. S., Holmes, N. C., Ross, M., Webb, D. J. & Pike, C. Shock temperatures and melting of iron at Earth core conditions. *Phys. Rev. Lett.* **70**, 3931–3934 (1993).
28. Yoo, C. S., Akella, J., Campbell, A. J., Mao, H. K. & Hemley, R. J. Phase diagram of iron by in situ X-ray diffraction: Implications for Earth’s core. *Science* **270**, 1473–1474 (1995).

Acknowledgements Discussions with P. Korzhavy were helpful. The calculations were done using the resources of the Swedish National Supercomputer Centre in Linköping. The study was supported by the Swedish Research Council (VR) and the Swedish Foundation for Strategic Research (SSF).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.B. (anatoly@fysik.uu.se).

Neutral theory and relative species abundance in ecology

Igor Volkov¹, Jayanth R. Banavar¹, Stephen P. Hubbell^{2,3} & Amos Maritan^{4,5}

¹Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

²Department of Plant Biology, The University of Georgia, Athens, Georgia 30602, USA

³The Smithsonian Tropical Research Institute, Box 2072, Balboa, Panama

⁴International School for Advanced Studies (SISSA), Via Beirut 2/4, 34014 Trieste, Italy

⁵INFN and The Abdus Salam International Center for Theoretical Physics, Trieste, Italy

The theory of island biogeography¹ asserts that an island or a local community approaches an equilibrium species richness as a result of the interplay between the immigration of species from the much larger metacommunity source area and local extinction of species on the island (local community). Hubbell² generalized this neutral theory to explore the expected steady-state distribution of relative species abundance (RSA) in the local community under restricted immigration. Here we present a theoretical framework for the unified neutral theory of biodiversity² and an analytical solution for the distribution of the RSA both in the metacommunity (Fisher's log series) and in the local community, where there are fewer rare species. Rare species are more extinction-prone, and once they go locally extinct, they take longer to re-immigrate than do common species. Contrary to recent assertions³, we show that the analytical solution provides a better fit, with fewer free parameters, to the RSA distribution of tree species on Barro Colorado Island, Panama⁴, than the lognormal distribution^{5,6}.

The neutral theory in ecology^{2,7} seeks to capture the influence of speciation, extinction, dispersal and ecological drift on the RSA under the assumption that all species are demographically alike on a per capita basis. This assumption, while only an approximation^{8–10}, appears to provide a useful description of an ecological community on some spatial and temporal scales^{2,7}. More significantly, it allows the development of a tractable null theory for testing hypotheses about community assembly rules. However, until now, there has been no analytical derivation of the expected equilibrium distribution of RSA in the local community, and fits to the theory have required simulations² with associated problems of convergence times, unspecified stopping rules, and precision³.

The dynamics of the population of a given species is governed by generalized birth and death events (including speciation, immigration and emigration). Let $b_{n,k}$ and $d_{n,k}$ represent the probabilities of birth and death, respectively, in the k th species with n individuals with $b_{-1,k} = d_{0,k} = 0$. Let $p_{n,k}(t)$ denote the probability that the k th species contains n individuals at time t . In the simplest scenario, the time evolution of $p_{n,k}(t)$ is regulated by the master equation^{11–13}

$$\frac{dp_{n,k}(t)}{dt} = p_{n+1,k}(t)d_{n+1,k} + p_{n-1,k}(t)b_{n-1,k} - p_{n,k}(t)(b_{n,k} + d_{n,k}) \quad (1)$$

which leads to the steady-state or equilibrium solution, denoted by P :

$$P_{n,k} = P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad (2)$$

for $n > 0$ and where $P_{0,k}$ can be deduced from the normalization condition $\sum_n P_{n,k} = 1$. Note that there is no requirement of

Box 1

Derivation of the RSA of the local community

We study the dynamics within a local community following the mathematical framework of McKane *et al.*²⁷, who studied a mean-field stochastic model for species-rich communities. In our context, the dynamical rules² governing the stochastic processes in the community are:

(1) With probability $1-m$, pick two individuals at random from the local community. If they belong to the same species, no action is taken. Otherwise, with equal probability, replace one of the individuals with the offspring of the other. In other words, the two individuals serve as candidates for death and parenthood.

(2) With probability m , pick one individual at random from the local community. Replace it by a new individual chosen with a probability proportional to the abundance of its species in the metacommunity. This corresponds to the death of the chosen individual in the local community followed by the arrival of an immigrant from the metacommunity. Note that the sole mechanism for replenishing species in the local community is immigration from the metacommunity, which for the purposes of local community dynamics is treated as a permanent source pool of species, as in the theory of island biogeography¹.

These rules are encapsulated in the following expressions for effective birth and death rates for the k th species:

$$b_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} + m \frac{\mu_k}{J_M} \left(1 - \frac{n}{J}\right) \quad (8)$$

$$d_{n,k} = (1-m) \frac{n}{J} \frac{J-n}{J-1} + m \left(1 - \frac{\mu_k}{J_M}\right) \frac{n}{J} \quad (9)$$

where μ_k is the abundance of the k th species in the metacommunity and J_M is the total population of the metacommunity.

The right hand side of equation (8) consists of two terms. The first corresponds to rule (1) with a birth in the k th species accompanied by a death elsewhere in the local community. The second term accounts for an increase of the population of the k th species due to immigration from the metacommunity. The immigration is, of course, proportional to the relative abundance μ_k/J_M of the k th species in the metacommunity. Equation (9) follows in a similar manner. Note that $b_{n,k}$ and $d_{n,k}$ not only depend on the species label k but also are no longer simply proportional to n .

Substituting equation (8) and (9) into equation (2), one obtains the expression²⁷:

$$P_{n,k} = \frac{J!}{n!(J-n)!} \frac{\Gamma(n+\lambda_k)}{\Gamma(\lambda_k)} \frac{\Gamma(\vartheta_k-n)}{\Gamma(\vartheta_k-J)} \frac{\Gamma(\lambda_k+\vartheta_k-J)}{\Gamma(\lambda_k+\vartheta_k)} \equiv F(\mu_k) \quad (10)$$

where

$$\lambda_k = \frac{m}{(1-m)}(J-1) \frac{\mu_k}{J_M} \quad (11)$$

and

$$\vartheta_k = J + \frac{m}{(1-m)}(J-1) \left(1 - \frac{\mu_k}{J_M}\right) \quad (12)$$

Note that the k dependence in equation (10) enters only through μ_k . On substituting equation (10) into equation (4), one obtains:

$$\langle \phi_n \rangle = \sum_{k=1}^{S_M} F(\mu_k) = S_M \langle F(\mu_k) \rangle = S_M \int d\mu \hat{\rho}(\mu) F(\mu) \quad (13)$$

Here $\hat{\rho}(\mu)d\mu$ is the probability distribution of the mean populations of the species in the metacommunity and has the form of the familiar Fisher log series (in a singularity-free description^{15,28}):

$$\hat{\rho}(\mu)d\mu = \frac{1}{\Gamma(\varepsilon)\delta^\varepsilon} \exp(-\mu/\delta) \mu^{\varepsilon-1} d\mu \quad (14)$$

where $\delta = x/(1-x)$. Substituting equation (14) into the integral in equation (13), taking the limits $S_M \rightarrow \infty$ and $\varepsilon \rightarrow 0$ with $\theta = S_M \varepsilon$ approaching a finite value^{15,28} and on defining $y = \mu_{\theta\delta}^{\frac{\theta}{\theta-1}}$, one can obtain our central result, equation (7).

conservation of community size. One can show that the system is guaranteed to reach the stationary solution (2) in the infinite time limit¹⁴.

The frequency of species containing n individuals is given by:

$$\phi_n = \sum_{k=1}^S I_k \quad (3)$$

where S is the total number of species and the indicator I_k is a random variable which takes the value 1 with probability $P_{n,k}$ and 0 with probability $(1 - P_{n,k})$. Thus the average number of species containing n individuals is given by:

$$\langle \phi_n \rangle = \sum_{k=1}^S P_{n,k} \quad (4)$$

The RSA relationship we seek to derive is the dependence of $\langle \phi_n \rangle$ on n .

Let a community consist of species with $b_{n,k} \equiv b_n$ and $d_{n,k} \equiv d_n$ being independent of k (the species are assumed to be demographically identical).

From equation (4), it follows that $\langle \phi_n \rangle$ is simply proportional to P_n , leading to:

$$\langle \phi_n \rangle = SP_0 \prod_{i=0}^{n-1} \frac{b_i}{d_{i+1}} \quad (5)$$

We consider a metacommunity in which the probability d that an individual dies and the probability b that an individual gives birth to an offspring are independent of the population of the species to which it belongs (density-independent case), that is, $b_n = bn$ and $d_n = dn$ ($n > 0$). Speciation may be introduced by ascribing a non-zero probability of the appearance of an individual of a new species, that is, $b_0 = v \neq 0$. Substituting the expressions into equation (5),

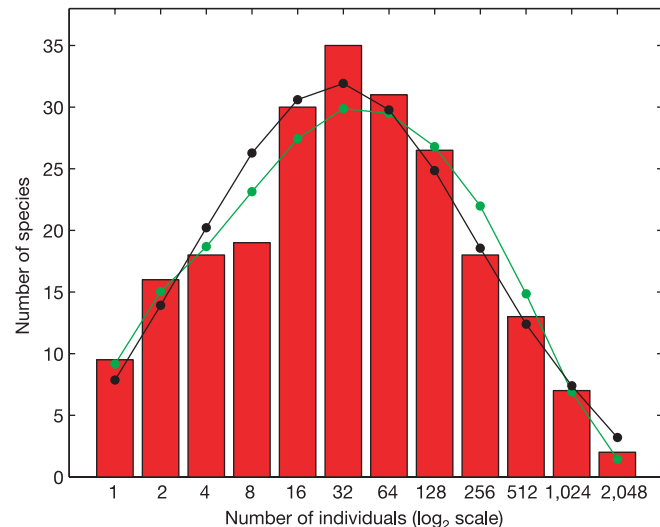


Figure 1 Data on tree species abundances in 50-hectare plot of tropical forest in Barro Colorado Island, Panama⁴. The total number of trees >10 cm DBH in the data set is 21,457 and the number of distinct species is 225. The red bars are observed numbers of species binned into log₂ abundance categories, following Preston's method⁵. The first histogram bar represents $\langle \phi_1 \rangle/2$, the second bar $\langle \phi_1 \rangle/2 + \langle \phi_2 \rangle/2$, the third bar $\langle \phi_2 \rangle/2 + \langle \phi_3 \rangle + \langle \phi_4 \rangle/2$ the fourth bar $\langle \phi_4 \rangle/2 + \langle \phi_5 \rangle + \langle \phi_6 \rangle + \langle \phi_7 \rangle + \langle \phi_8 \rangle/2$ and so on. The black curve shows the best fit to a lognormal distribution $\langle \phi_n \rangle = \frac{N}{n} \exp(-(\log_2 n - \log_2 n_0)^2 / 2\sigma^2)$ ($N = 46.29$, $n_0 = 20.82$ and $\sigma = 2.98$), while the green curve is the best fit to our analytic expression equation (7) ($m = 0.1$ from which one obtains $\theta = 47.226$ compared to the Hubbell² estimates of 0.1 and 50 respectively and McGill's best fits³ of 0.079 and 48.5 respectively).

one obtains the celebrated Fisher log series¹⁵:

$$\langle \phi_n^M \rangle = S_M P_0 \frac{b_0 b_1 \dots b_{n-1}}{d_1 d_2 \dots d_n} = \theta \frac{x^n}{n} \quad (6)$$

where M refers to the metacommunity, $x = b/d$ and $\theta = S_M P_0 v/b$ is the biodiversity parameter (also called Fisher's α). We follow the notation of Hubbell² in this paper. Note that x represents the ratio of effective per capita birth rate to the death rate arising from a variety of causes such as birth, death, immigration and emigration. Note that in the absence of speciation, $b_0 = v = \theta = 0$, and, in equilibrium, there are no individuals in the metacommunity. When one introduces speciation, x has to be less than 1 to maintain a finite metacommunity size $J_M = \sum_n n \langle \phi_n \rangle = \theta x / (1 - x)$.

We turn now to the case of a local community of size J undergoing births and deaths accompanied by a steady immigration of individuals from the surrounding metacommunity. When the local community is semi-isolated from the metacommunity, one may introduce an immigration rate m , which is the probability of immigration from the metacommunity to the local community. For constant m (independent of species), immigrants belonging to the more abundant species in the metacommunity will arrive in the local community more frequently than those of rarer species.

Our central result (see Box 1 for a derivation) is an analytic expression for the RSA of the local community:

$$\langle \phi_n \rangle = \theta \frac{J!}{n!(J-n)!} \frac{\Gamma(\gamma)}{\Gamma(J+\gamma)} \int_0^1 \frac{\Gamma(n+y)}{\Gamma(1+y)} \frac{\Gamma(J-n+\gamma-y)}{\Gamma(\gamma-y)} \exp(-y\theta/\gamma) dy \quad (7)$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ which is equal to $(z-1)!$ for integer z and $\gamma = \frac{m(J-1)}{1-m}$. As expected, $\langle \phi_n \rangle$ is zero when n exceeds J . The computer calculations in Hubbell's book² as well as those more recently carried out by McGill³ were aimed at estimating $\langle \phi_n \rangle$ by simulating the processes of birth, death and immigration.

One can evaluate the integral in equation (7) numerically for a given set of parameters: J , θ and m . For large values of n , the integral can be evaluated very accurately and efficiently using the method of steepest descent¹⁶. Any given RSA data set contains information about the local community size, J , and the total number of species in the local community, $S_L = \sum_{k=1}^J \langle \phi_k \rangle$. Thus there is just one free fitting parameter at one's disposal.

McGill asserted³ that the lognormal distribution is a more parsimonious null hypothesis than the neutral theory, a suggestion which is not borne out by our reanalysis of the Barro Colorado Island (BCI) data. We focus only on the BCI data set because, as pointed out by McGill³, the North American Breeding Bird Survey data are not as exhaustively sampled as the BCI data set, resulting in fewer individuals and species in any given year in a given location. Furthermore, the McGill analysis seems to rely on adding the bird counts over five years at the same sampling locations even though these data sets are not independent.

Figure 1 shows a Preston-like binning⁵ of the BCI data⁴ and the fit of our analytic expression with one free parameter (11 degrees of freedom) along with a lognormal having three free parameters (9 degrees of freedom). Standard chi-square analysis¹⁷ yields values of $\chi^2 = 3.20$ for the neutral theory and 3.89 for the lognormal. The probabilities of such good agreement arising by chance are 1.23% and 8.14% for the neutral theory and lognormal fits, respectively. Thus one obtains a better fit of the data with the analytical solution to the neutral theory to BCI than with the lognormal, even though there are two fewer free parameters. McGill's analysis³ on the BCI data set was based on computer simulations in which there were difficulties in knowing when to stop the simulations, that is, when equilibrium had been reached. It is unclear whether McGill averaged over an ensemble of runs, which is essential to obtain repeatable and reliable results from simulations of stochastic

processes because of their inherent noisiness. However, simulations of the neutral theory are no longer necessary, and all problems with simulations are moot, because an analytical solution is now available.

The lognormal distribution is biologically less informative and mathematically less acceptable as a dynamical null hypothesis for the distribution of RSA than the neutral theory. The parameters of the neutral theory or RSA are directly interpretable in terms of birth and death rates, immigration rates, size of the metacommunity, and speciation rates. A dynamical model of a community cannot yield a lognormal distribution with finite variance because in its time evolution, the variance increases through time without bound. However, as shown in ref. 18, the lognormal distribution can arise in static models, such as those based on niche hierarchy.

The steady-state deficit in the number of rare species compared to that expected under the log series can also occur because rare species grow differentially faster than common species and therefore move up and out of the rarest abundance categories owing to their rare-species advantage¹⁹. Indeed, it is likely that several different models (such as an empirical lognormal distribution, niche hierarchy models¹⁸ or the theory presented here) might provide comparable fits to the RSA data (we have found that the lognormal does slightly better than the neutral theory for the Pasoh data set²⁰, obtained in a tropical tree community in Malaysia). Such fitting exercises in and of themselves, however, do not constitute an adequate test of the underlying theory. Neutral theory predicts that the degree of skewing of the RSA distribution ought to increase as the rate of immigration into the local community decreases. Dynamic data on rates of birth, death, dispersal and immigration are needed to evaluate the assumptions of neutral theory and determine the role played by niche differentiation in the assembly of ecological communities.

Our analysis should also apply to the field of population genetics in which the mutation-extinction equilibrium of neutral allele frequencies at a given locus has been studied for several decades^{21–26}. □

Received 7 May; accepted 9 June 2003; doi:10.1038/nature01883.

- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, NJ, 1967).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
- McGill, B. J. A test of the unified neutral theory of biodiversity. *Nature* **422**, 881–885 (2003).
- Condit, R. *et al.* Beta-diversity in tropical forest trees. *Science* **295**, 666–669 (2002).
- Preston, F. W. The commonness and rarity of species. *Ecology* **29**, 254–283 (1948).
- May, R. M. *Ecology and Evolution of Communities* 81–120 (Harvard Univ. Press, Cambridge, MA, 1975).
- Bell, G. Neutral macroecology. *Science* **293**, 2413–2418 (2001).
- Diamond, J. & Case, T. J. (eds) *Community Ecology* (Harper and Row, New York, NY, 1986).
- Tilman, D. *Plant Strategies and the Dynamics and Structure of Plant Communities* (Princeton Univ. Press, Princeton, NJ, 1988).
- Weiner, E. & Keddy, P. *Ecological Assembly Rules: Perspectives, Advances, and Retreats* (Cambridge Univ. Press, Cambridge, UK, 1999).
- Boswell, M. T., Ord, J. K. & Patil, G. P. *Statistical Distributions in Ecological Work* 3–157 (International Co-operative Publishing, Fairland, MD, 1979).
- Caraco, T. *Statistical Distributions in Ecological Work* 371–387 (International Co-operative Publishing, Fairland, MD, 1979).
- Feller, W. *An Introduction to Probability Theory and Its Applications* Vol. 1 (John Wiley & Sons, Hoboken, NJ, 1968).
- Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (Amsterdam, North-Holland, 2001).
- Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58 (1943).
- Morse, P. M. & Feshbach, H. *Methods of Theoretical Physics* Part 1 (McGraw-Hill, New York, NY, 1953).
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, UK, 1993).
- Sugihara, G., Bersier, L., Southwood, T. R. E., Pimm, S. L. & May, R. M. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA* **100**, 5246–5251 (2003).
- Chave, J., Muller-Landau, H. C. & Levin, S. A. Comparing classical community models: Theoretical consequences for patterns of diversity. *Am. Nat.* **159**, 1–23 (2002).
- Manokaran, N. *et al.* *Stand Tables and Species Distributions in the Fifty Hectare Plot at Pasoh Forest Reserve* (Forest Research Institute Malaysia, Kuala Lumpur, 1992).
- Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
- Ewens, W. J. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**, 87–112 (1972).
- Karlin, J. & McGregor, J. Addendum to a paper of W. Ewens. *Theor. Pop. Biol.* **3**, 113–116 (1972).

- Watterson, G. A. Models for the logarithmic species abundance distributions. *Theor. Pop. Biol.* **6**, 217–250 (1975).
- Kimura, M. & Ohta, T. *Theoretical Aspects of Population Genetics* (Princeton Univ. Press, Princeton, NJ, 1971).
- Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, UK, 1983).
- McKane, A., Alonso, D. & Solé, R. V. Mean-field stochastic theory for species-rich assembled communities. *Phys. Rev. E* **62**, 8466–8484 (2000).
- Rao, C. R. *Statistical Ecology* Vol. 1 *Spatial Patterns and Statistical Distributions* 131–142 (The Penn. State Univ. Press, University Park, PA, 1971).

Acknowledgements We are grateful to O. Kargaltsev for a careful reading of the manuscript. This work was supported by NASA, by grants from the NSF, and by the Department of Plant Biology, University of Georgia.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.B. (banavar@psu.edu) or S.P.H. (shubbell@dogwood.botany.uga.edu).

The genome of a motile marine *Synechococcus*

B. Palenik¹, B. Brahamsha¹, F. W. Larimer^{2,3}, M. Land^{2,3}, L. Hauser^{2,3}, P. Chain^{3,4}, J. Lamerdin^{3,4}, W. Regala^{3,4}, E. E. Allen^{1*}, J. McCarren¹, I. Paulsen⁵, A. Dufresne⁶, F. Partensky⁶, E. A. Webb⁷ & J. Waterbury⁷

¹Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0202, USA

²Computational Biology, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6480, USA

³Joint Genome Institute, Walnut Creek, California 94598, USA

⁴Lawrence Livermore National Laboratory, Livermore, California 94550-9234, USA

⁵TIGR, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

⁶UMR 7127 CNRS Station Biologique de Roscoff, 29682 Roscoff, France

⁷Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

* Present address: Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA

Marine unicellular cyanobacteria are responsible for an estimated 20–40% of chlorophyll biomass and carbon fixation in the oceans¹. Here we have sequenced and analysed the 2.4-megabase genome of *Synechococcus* sp. strain WH8102, revealing some of the ways that these organisms have adapted to their largely oligotrophic environment. WH8102 uses organic nitrogen and phosphorus sources and more sodium-dependent transporters than a model freshwater cyanobacterium. Furthermore, it seems to have adopted strategies for conserving limited iron stores by using nickel and cobalt in some enzymes, has reduced its regulatory machinery (consistent with the fact that the open ocean constitutes a far more constant and buffered environment than fresh water), and has evolved a unique type of swimming motility. The genome of WH8102 seems to have been greatly influenced by horizontal gene transfer, partially through phages. The genetic material contributed by horizontal gene transfer includes genes involved in the modification of the cell surface and in swimming motility. On the basis of its genome, WH8102 is more of a generalist than two related marine cyanobacteria².

Most species of picoplanktonic marine cyanobacteria currently known belong to two genera: *Synechococcus* and *Prochlorococcus*. Members must have the ability to acquire major nutrients and trace metals at the submicromolar concentrations found in the oligotrophic open seas. Their light-harvesting apparatus is uniquely

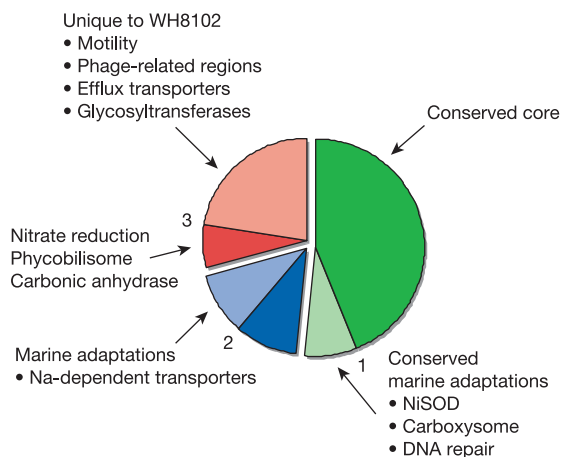


Figure 1 The genome of WH8102 can be divided into three categories: ORFs found in WH8102 and both related *Prochlorococcus* genomes (region 1); ORFs found in WH8102 and only one *Prochlorococcus* genome (region 2); and ORFs not found in the two *Prochlorococcus* genomes (region 3). These regions can be subdivided based on whether or not an ORF has a BLAST hit to freshwater *Synechocystis* sp. strain PCC6803. The lighter shading of each region represents the fraction not in PCC6803. As shown, examination of ORFs in each subcategory has provided insights into the evolutionary adaptation of marine *Synechococcus*.

adapted to the spectral quality of light in the ocean^{3,4}. Of the two major marine unicellular genera, *Synechococcus* is usually less abundant in very oligotrophic environments, but has a broader global distribution^{1,4}. A great deal of genetic diversity exists within the genus *Synechococcus*, with strains probably adapted to specific ecological niches^{3,5}. Furthermore, one group of strains seems to have adapted to the oligotrophic marine environment by developing a new form of swimming motility not seen so far in any other prokaryotic group⁶.

On the basis of the genome of *Synechococcus* sp. strain WH8102 (hereafter referred to as WH8102) and comparing it to the related genomes of two *Prochlorococcus* strains² we were able to define 1,314 open reading frames (ORFs) common to all three genomes (about half of the genome) and 736 ORFs found only in WH8102 (about a third of the genome) that indicate the specific ecological strategies of WH8102 relative to coexisting *Prochlorococcus* (see Methods, Fig. 1 and Supplementary Table 1). Here we show how the genome reveals some of the interactions of WH8102 with its environment

(nutrients, light, toxins) and with other organisms, especially phages.

The WH8102 genome contains 16 probable or possible phage integrases—enzymes that function as site-specific DNA recombinases⁷ (Supplementary Table 2). In WH8102, many of these occur adjacent to or near transfer RNAs and in regions with an anomalously low percentage of G + C content (Table 1a and Fig. 2). These regions of low G + C percentage also show atypical trinucleotide composition (data not shown). In addition, possible phage integrase regulators (SYNW2105, SYNW1660, SYNW1665) are also found. Thus, *Synechococcus* has regions that greatly resemble pathogenicity islands—regions that are often mobilized between strains of pathogenic bacteria⁷. Hence, although the genome of WH8102 does not contain prophages or plasmids, it does seem to have been, in its evolutionary past, extensively altered through horizontal gene transfer, possibly due to phages or plasmids. In contrast, far fewer potential phage integrases are found in *Prochlorococcus* (four in MIT9313 and one in MED4).

Other regions of low G + C percentage not associated with phage integrases also show atypical trinucleotide composition, suggestive of recent horizontal gene transfer (Table 1b). These regions encode genes involved in broadening the range of nitrogen substrates that can be used by WH8102, as well as some encoding transport capabilities. Furthermore, several of the genes found in these regions are homologues of ORFs involved in the carbohydrate modification of the cell envelope (including glycosyltransferases, and homologues of genes involved in the synthesis of sialic acid). One hypothesis for the function of these glycosyltransferases is that they may be required in constructing the motility apparatus of this organism, as at least one of its components is glycosylated⁸. Another possibility is that WH8102 may use these envelope-modifying genes to change its cell surface characteristics to help it evade grazers and other predators such as phages. Cell surface properties are known to affect grazing rates in the marine environment⁹.

An examination of the genome of WH8102 provides an indication of the uniqueness of *Synechococcus* swimming motility. None of the proteins (motor, flagellar) associated with other forms of prokaryotic motility was found, with the exception of six ORFs associated with type IV-pilus-dependent motility (homologues of *pilB*, *-C*, *-D*, *-Q* and *-T*). Orthologues of these are also present in MIT9313, but not MED4. Nevertheless, these ORFs in WH8102 do not encode the full complement of genes required for pilus assembly and function, and pilin subunit homologues are absent. Pili or surface-associated twitching have not been observed in WH8102.

Recent studies using transposon mutagenesis (J.M. and B.B., manuscript in preparation) coupled with that of motility mutant *swmA*⁸ indicate that genes required for motility are found in at least

Table 1 Atypical regions of G + C per cent in the WH8102 genome

Region	Size (kb)	No. ORFs	G + C (%)	Acquisition
(a) Low G + C per cent regions associated with predicted phage integrases				
1124989–1158432	33.4	27	53.6	–
1982294–1996886	14.6	14	50.7	Efflux
1606831–1591392	15.4	12	47.8	–
1335428–1344827	9.4	7	49.9	–
2312441–2322542	10.1	12	50.7	–
1488335–1524808	36.0	37	49.8	–
1183175–1171388	11.8	13	51.7	–
857915–842611	15.3	14	49.3	Na/Glut. symporter
2138960–2144868	5.9	4	40.9	–
353145–383689	30.0	29	52.0	–
(b) Low G + C per cent regions not associated with predicted phage integrases				
427233–465883	38.7	37	38.8	Modification of cell envelope, multidrug efflux
622199–633146	10.9	7	43.96	Modification of cell envelope
2379778–2394189	14.4	15	39.0	Cyanate usage, metal uptake
912098–954990	42.9	12	42.2	Motility

two widely separated regions (Fig. 3). The second region contains SwmB (SYNW0953), a very large ORF (10,791 amino acids) that constitutes more than 1% of the genome size and is currently one of the longest bacterial ORFs ever reported. Notably, it is found in one of the unusually low G + C percentage regions (Table 1b).

Transport in WH8102 accounts for about 5–6% of the predicted ORFs, similar to most other bacterial genomes¹⁰. Compared with other genomes¹⁰, transporter capability is heavily biased towards the use of ABC transporters with about 60% of the ORFs encoding ABC transporter components. A distinct bias against P-type ATPase transporters is found, with only one such transporter, for copper, compared with nine in PCC6803, a model unicellular freshwater cyanobacterium. This one P-type ATPase may be conserved due to the use of copper in plastocyanin, an electron transfer protein that can substitute for an iron-containing cytochrome in photosynthesis.

Notably, WH8102 has multiple channels for transporting major seawater ions and multiple transporters or ABC-type solute-

binding proteins for several major nutrients; for example, there are multiple solute-binding proteins for phosphate and two for urea. WH8102 seems to have an independent transporter for urea (SYNW2455), reinforcing its importance as a nitrogen source for cyanobacterial growth in oligotrophic environments. The multiple transporters in WH8102 may have different affinities and be regulated differently depending on nutrient concentrations.

One of the surprises from our analyses of the genome of WH8102 is the prediction that *Synechococcus* can use some new organic compounds as nitrogen and phosphorus sources. As inorganic nitrogen and phosphorus are often thought to be limiting in the marine environment, these potential sources are of particular interest. Amino acid and oligopeptide transporters are found, suggesting that *Synechococcus* may have the ability to use these ubiquitous compounds in sea water; transport of a few amino acids has also been demonstrated¹¹. In addition, genes for the transport of cyanate and its breakdown by cyanase appear to be present in WH8102 and in *Prochlorococcus* MED4—in fact, WH8102 grows on

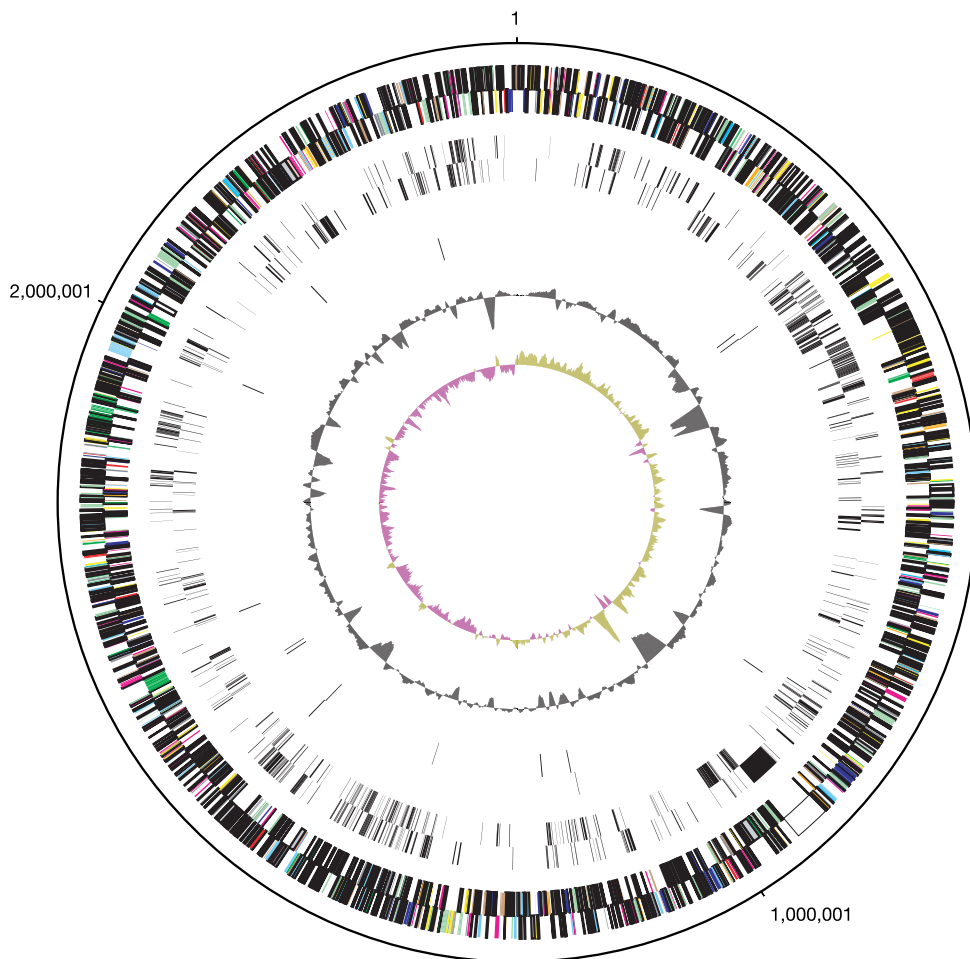


Figure 2 The chromosome of *Synechococcus* sp. strain WH8102. The following description is of the eight circles, with the first circle being the outer circle, and the others progressing inwards. First circle, predicted coding regions on the plus strand coloured by functional category: white, hypothetical; black, unassigned and other; dark red, energy metabolism; green, photosynthesis; blue, DNA replication and repair; cyan, fatty acid metabolism; magenta, biosynthesis of cofactors; yellow, cellular processes; pale green, transport and binding; sky blue, translation; orange, regulatory functions; brown, amino acid biosynthesis; pink, cell envelope; grey, conserved hypothetical; medium red, transcription; light red, purines and pyrimidines; pale pink, central metabolism. Second

circle, predicted coding regions on minus strand (same colour scheme as in the first circle). Third and fourth circles, 736 'characteristic' genes not found in two *Prochlorococcus* strains (region 3 in Fig. 1), plus and minus strands, respectively. Fifth and sixth circles, predicted phage integrases on plus and minus strands, respectively. Seventh circle, G + C content (deviation from average); eighth circle, G + C skew curve in purple and olive. Phage integrases are often associated with low G + C regions. Low G + C regions often contain WH8102 characteristic ORFs. Scale (in bp) is indicated along the outside of the circle.

cyanate as a sole nitrogen source (B.P., unpublished data). The genes for cyanate use have been characterized for the freshwater cyanobacterium *Synechococcus elongatus* PCC7942 but cyanate remains uncharacterized as a nitrogen source in aquatic environments¹².

Similarly, genes for the transport of phosphonates (compounds with C–P bonds) are present in WH8102 and are found in the two sequenced *Prochlorococcus* genomes as well. WH8102 grows on phosphonate as a sole phosphorous source (B.P., unpublished data). Phosphonates are known to be produced by some major eukaryotic phytoplankton groups such as the coccolithophorids, and were recently reported to be an important fraction of total phosphate in sea water¹³. WH8102 also has multiple alkaline phosphatases (SYNW0120, SYNW0196, SYNW2391 and SYNW2390) that could be used to obtain phosphate from other organic phosphorus sources in its environment. Thus genome analyses are further dispelling the classical concept of cyanobacteria as being plant-like and dependent solely on inorganic forms of nutrients¹⁴.

In addition, a number of conserved systems for exporting compounds (for example, multidrug efflux systems) are found both in the ABC transporter family and the MFS transporter family. WH8102 has a larger number of efflux transporters in the ABC family compared with *Prochlorococcus*. These results suggest that marine cyanobacteria, despite living in extremely oligotrophic conditions, may still find themselves in the position of needing to export ‘toxins’ produced by other microorganisms. Antagonistic interactions between pelagic bacteria have been reported recently¹⁵. Exposure to toxins may be greater for motile *Synechococcus* than for other marine cyanobacteria, as they may be chemotactic towards marine particles where higher localized concentrations of heterotrophic bacteria release nitrogenous compounds¹¹.

WH8102 also has efflux pumps for metals (SYNW1472 and SYNW0900) that are lacking in both *Prochlorococcus* strains. Characterizing these further may put a mechanistic basis behind previous observations¹⁶ that *Synechococcus* seems to be more resistant

to copper compared with *Prochlorococcus*, and that this resistance may help explain the seasonal cycles of these organisms in the Sargasso Sea. WH8102 also has predicted genes for the reduction of arsenate to arsenite (SYNW1767) and its efflux (SYNW1039). It has been hypothesized that arsenate is a competitor for phosphate and that systems would be needed to deal with this compound, especially in low-phosphate waters¹⁷.

WH8102 has more capacity for sodium-driven transport than freshwater cyanobacteria such as PCC6803 (see Methods and Fig. 1), with transporters of the alanine/glycine:cation (sodium) symporter family (SYNW0828) and of the neurotransmitter:sodium symporter family (SYNW0699). It also has two transporters from the solute:sodium symporter family (SYNW2455, SYNW0619) compared with one in PCC6803.

In contrast to freshwater cyanobacteria, WH8102 has two potential transporters (SYNW1915, SYNW1916 and SYNW1917, and SYNW0229) for glycine betaine and related compounds found in marine waters. Adjacent to the ABC transporter but on the opposite strand are enzymes predicted to synthesize glycine betaine from glycine (SYNW1914, SYNW1913) using a pathway only reported before from an extremely halophilic proteobacterium¹⁸. When a freshwater *Synechococcus*, strain PCC7942, was genetically engineered to make glycine betaine, it became more halotolerant¹⁹.

Despite the importance of iron as a limiting nutrient in the oceans, WH8102 does not have a detectable system for siderophore synthesis and uptake. However, it does have strategies for iron conservation such as using plastocyanin (copper) for electron transport and a cobalt-dependent ribonucleotide reductase (SYNW1692) rather than the Fe-containing one found in many other cyanobacteria. Another example of iron conservation in WH8102 is its predicted nickel superoxide dismutase (SOD). Multiple SOD types exist for removing photosynthetically produced superoxide radicals including ones using iron, manganese or copper-zinc as metal cofactors. Unlike the freshwater PCC6803, the marine cyanobacteria WH8102, both *Prochlorococcus* species and *Trichodesmium*, a marine N₂-fixing cyanobacterium (http://www.jgi.doe.gov/JGI_microbial/html/index.html), are predicted to use a new nickel SOD—seen recently in *Streptomyces*—as their only SOD, thus saving iron and manganese for other uses (see Supplementary Fig. 1).

In comparison with the sequenced *Prochlorococcus* strains, WH8102 is a transport generalist. It has predicted transporters for the efflux of chromate (SYNW1323) and arsenite (SYNW1039) that are found in MED4 but not MIT9313. It shares with MIT9313, but not MED4, the ability to use the sodium symporters mentioned above. In addition, WH8102 has transporters that are not found in *Prochlorococcus*, and that are predicted to be involved in the uptake of nitrate, a quaternary ammonium group (R–N+(CH₃)₃) compound such as sarcosine, another nitrate-like compound, metals (magnesium/cobalt/nickel), and in cation efflux. This may be a characteristic of marine *Synechococcus* in general or it may be a characteristic of motile *Synechococcus*.

In WH8102 the major components of photosynthesis and respiration are well conserved and are usually most closely related to those of other cyanobacteria. Notable exceptions are genes in WH8102 and *Prochlorococcus* implicated in carboxysome structure and assembly, including those encoding the subunits of ribulose-1,5-bisphosphate carboxylase. This is thought to be due to a horizontal gene transfer event²⁰.

As in most cyanobacteria (other than marine *Prochlorococcus* and two other known prochlorophytes), *Synechococcus* harvests light using phycobilisomes, which are multisubunit complexes binding different types of phycobilins. Two interesting observations differentiate the use of phycobilisomes by WH8102 from those of other cyanobacteria analysed so far. Nowhere in the WH8102 genome are there homologues of the *cpcC* and *cpcD* genes, which in freshwater cyanobacteria are known to encode two types of phycocyanin-

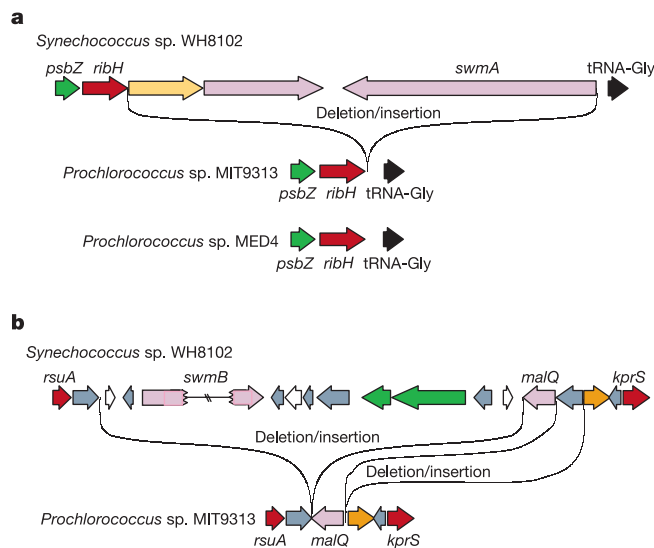


Figure 3 Organization of two chromosomal regions in WH8102 that contain motility genes. **a**, The region containing *swmA* and two other ORFs (a predicted glycosyltransferase and a sulphotransferase) appears to have been inserted in the *ribH*/tRNA-Gly region, or conversely, deleted in the two *Prochlorococcus* genomes. **b**, *swmB* and several other ORFs have been inserted between the *rsuA* and *malQ* regions. The double lines in *swmB* indicate that it is not drawn to scale, as it is approximately 20 times larger than *malQ*. The ORFs are colour-coded by predicted function as described in Fig. 2.

associated L_R linker polypeptides. These linkers are necessary for the correct assembly of phycocyanin discs in the phycobilisome rods²¹. Their absence in WH8102 suggests that there is a single disc of phycocyanin, as is the case in mutants of *Synechococcus* PCC7002 in which the *cpcC* gene has been inactivated²¹. The genome thus provides a basis for the interpretation of absorbance spectra, where reduced phycocyanin (orange-light absorbing) relative to phycoerythrin (blue-light absorbing) probably represents an adaptation to the oligotrophic marine environment where blue light is particularly important³. In addition, the genome of WH8102 also lacks homologues of *nblA* and *nblB*, two genes implicated in the degradation of phycobilisomes during nutrient stress in cyanobacteria^{22,23}. Thus, phycobilisome degradation may not occur or may be under the control of other genes in WH8102.

Whereas *Synechococcus* possesses homologues of the low-affinity bicarbonate transport mechanism in PCC6803, it lacks homologues of *ndhD3*, *ndhF3* and *chpY*, genes implicated in high-affinity transport in the same organism²⁰. Their absence might be an adaptation to the marine habitat, where bicarbonate (2 mM) is probably rarely limiting. Notably, WH8102 has two predicted carbonic anhydrases (SYNW0897 and SYNW2467) whereas *Prochlorococcus* has none, although these genes can be highly divergent and difficult to predict. SYNW2467 is adjacent to the genes encoding nitrate reductase. *Synechococcus* WH8102 can use nitrate for growth in contrast to the two *Prochlorococcus* strains², and this carbonic anhydrase may have been lost with the loss of nitrate usage. Although intriguing, a specific connection between nitrate usage and carbonic anhydrase has not been shown.

One way that bacteria sense and respond to their environment is by using two-component regulatory systems consisting of a sensor kinase and a response regulator. In PCC6803 there are nearly 40 sensor kinase and response regulator pairs (<http://www.kazusa.or.jp/cyano/index.html>). In contrast, WH8102 has only five sensor histidine kinases and nine response regulators, of which one, SYNW1598, may be a pseudogene as it is missing conserved functional residues²⁴. Even accounting for a smaller genome size, WH8102 as well as the two *Prochlorococcus* species have fewer systems for responding to environmental changes using these gene families. Furthermore, as there are fewer sensors than response regulators, there seems to be an economy of regulation in which some sensors may transmit information to more than one response regulator.

In addition to the principal RNA polymerase sigma factor *sigA* (SYNW1783), WH8102 encodes five type II sigma factors, typical of cyanobacteria in general²⁵. WH8102 however has only one homologue of the type III sigma factor (SYNW1232). This is a low number compared with the three to five seen in other sequenced cyanobacteria (PCC6803, PCC7120 and *Thermosynechococcus*; <http://www.kazusa.or.jp/cyano/index.html>). One hypothesis for the minimal regulatory machinery (two-component systems and sigma factors) in *Synechococcus* and *Prochlorococcus* is that they have evolved in an open ocean environment that is relatively constant, thus they do not need a regulatory system that could modulate their gene expression to a more variable environment. Alternatively, a minimal regulatory system could be the result of an ecological strategy of only some marine cyanobacteria.

On the basis of its genome, *Synechococcus* WH8102 is clearly more nutritionally versatile and a 'generalist' compared with its *Prochlorococcus* relatives. As the genus *Prochlorococcus* seems to have evolved only once, it may have gone through an evolutionary 'bottleneck' in which its capabilities were originally limited to those of a particular strain followed by subsequent acquisition of new abilities. Alternatively, *Synechococcus* may be more subject than *Prochlorococcus* to horizontal gene transfer from phages, as seen by the presence of more phage integrases. It is possible that not all *Synechococcus* are more versatile in their transport abilities, just the strains that are motile. Partial or complete genomes of additional

marine cyanobacteria from this group will help answer these questions. □

Methods

Genome sequencing

Genomic DNA was isolated from WH8102 as reported previously²⁶. Whole-genome shotgun libraries were obtained by fragmenting genomic DNA using mechanical shearing and cloning 2–3-kilobase fragments into pUC18. Double-ended plasmid sequencing reactions were carried out using PE BigDye Terminator chemistry (Perkin Elmer) and sequencing ladders were resolved on PE 377 Automated DNA Sequencers (Perkin Elmer). As the first genome drafted during the start-up of the microbial sequencing effort at the J.G.I. Production Sequencing Facility in Walnut Creek, California, this genome was sequenced to unusually high coverage. The whole-genome sequence of WH8102 was obtained from 66,550 reads with an average read length for this project of >575 base pairs (bp) per read for 16-fold redundancy. Sequence assembly was accomplished using PHRAP (P. Green). All gaps were closed by primer walking on gap-spanning library clones or PCR products. The overall genome structure was verified by long-range genomic PCR reactions. The two tandem repeats were resolved by combining information from individual clones, single-nucleotide polymorphism analysis and PCR. Only after this region was finished was it discovered that a single, long ORF was preserved.

Genome analysis

For genome analyses, the combination of three gene modelling programs—Critica²⁷, Glimmer²⁸ and Generation (<http://compbio.ornl.gov/generation/index.shtml>)—was used in the determination of potential coding sequences. These assignments were further checked manually. A revised gene/protein set was searched against the KEGG GENES, Pfam, PROSITE, PRINTS, ProDom and COGs databases, in addition to BLASTP against the NCBI non-redundant database. From these results, categorizations were developed using the KEGG and COGs hierarchies. Transfer RNAs were identified using tRNAscan-SE²⁹. To identify regions of atypical nucleotide composition, the trinucleotide composition was determined.

Manual annotation of ORFs was carried out using Artemis, the Artemis Comparison Tool (<http://www.sanger.ac.uk/Software/ACT/>) and Clustal W³⁰. The results of the KEGG and other comparisons described above were examined manually to check automated product assignments and make additional assignments. The proteome sequences of WH8102 and *Prochlorococcus* (MED4 and MIT9313)² were compared using the Artemis Comparison Tool. This program, in conjunction with Clustal W, was used for refining predicted start sites, adding ORFs not predicted by the gene modelling programs, and obtaining consistent annotation across three genomes. Manual annotation was done in conjunction with the *Prochlorococcus* annotation team. Transporters were analysed and annotated using methods described in ref. 10.

Pairwise BLAST analyses of three marine cyanobacterial genomes (WH8102 and *Prochlorococcus marinus* strains MED4 and MIT9313 (ref. 2)) against each other and a cut-off *e*-value of e^{-6} , followed by additional manual curation including examination of gene context, were used to partition the genome of WH8102 into three categories: 1,314 ORFs found in all three genomes and predicted to be orthologues; 476 predicted orthologous ORFs found in WH8102 and one other *Prochlorococcus* genome; and 736 ORFs characteristic of WH8102 (not found in either of the other *Prochlorococcus* genomes). The latter category partially represents the ecological capabilities of this organism compared with *Prochlorococcus* (Supplementary Table 3).

Using pairwise BLAST analyses, the three categories of WH8102 ORFs were further subdivided based on whether or not an ORF was found in a model freshwater cyanobacterium *Synechocystis* PCC6803 (hereafter termed as PCC6803; see <http://www.kazusa.or.jp/cyano/index.html>). After examination of different cut-offs, BLAST analyses with a cut-off *e*-value of e^{-10} were used for this assignment. For the 'core' marine cyanobacterial genome of 1,314 ORFs, 1,112 (85%) are also found in PCC6803 (Fig. 1). This provides an estimate of the portion of the WH8102 genome that has been conserved in all cyanobacterial genomes so far from a primal cyanobacterial ancestor (and includes ORFs conserved in all bacterial taxa). This portion is different from a minimal bacterial genome or a minimal cyanobacterial genome, as horizontally acquired genes could carry out functions required for cell viability. The 15% of the marine cyanobacterial core not in PCC6803 was found to include some of the adaptations and evolutionary events that distinguish the marine *Synechococcus*/*Prochlorococcus* cyanobacterial lineage from other cyanobacterial lineages.

Of the 736 WH8102 characteristic ORFs not found in the *Prochlorococcus* genomes, 23% have related ORFs in PCC6803, using a BLAST cut-off *e*-value of e^{-10} . This is not surprising as these partly represent the ability of both PCC6803 and WH8102, but not the *Prochlorococcus* strains, to create a functional phycobilisome for harvesting light, and a functional nitrate reductase with molybdenum cofactors for using nitrate as a nitrogen source. Forty-five per cent of these characteristic ORFs are hypothetical.

Received 9 May; accepted 28 July 2003; doi:10.1038/nature01943.

Published online 13 August 2003.

- Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
- Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047 (2003).
- Waterbury, J. B., Watson, F. W., Valois, F. W. & Franks, D. G. in *Photosynthetic Picoplankton* (eds Platt, T. & Li, W. K. W.) 71–120 (Canadian Department of Fisheries and Oceans, Ottawa, 1986).

4. Scanlan, D. J. & West, N. J. Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol. Ecol.* **40**, 1–12 (2002).
5. Ferris, M. J. & Palenik, B. Niche adaptation in ocean cyanobacteria. *Nature* **396**, 226–228 (1998).
6. Toledo, G., Palenik, B. & Brahamsha, B. Swimming strains of marine *Synechococcus* with widely different photosynthetic pigment ratios form a monophyletic group. *Appl. Environ. Microbiol.* **65**, 5247–5251 (1999).
7. Bushman, F. *Lateral DNA Transfer Mechanisms and Consequences* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
8. Brahamsha, B. An abundant cell-surface polypeptide is required for swimming by the nonflagellated marine cyanobacterium *Synechococcus*. *Proc. Natl Acad. Sci. USA* **93**, 6504–6509 (1996).
9. Monger, B. C., Landry, M. R. & Brown, S. L. Feeding selection of heterotrophic marine nanoflagellates based on the surface hydrophobicity of their picoplankton prey. *Limnol. Oceanogr.* **44**, 1917–1927 (1999).
10. Paulsen, I. T., Nguyen, L., Sliwinski, M. K., Rabus, R. & Saier, M. H. Jr Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* **301**, 75–101 (2000).
11. Willey, J. M. & Waterbury, J. B. Chemotaxis toward nitrogenous compounds by swimming strains of marine *Synechococcus* spp. *Appl. Environ. Microbiol.* **55**, 1888–1894 (1989).
12. Harano, Y. et al. Identification and nitrogen regulation of the cyanase gene from the cyanobacteria *Synechocystis* sp. strain PCC6803 and *Synechococcus* sp. strain PCC7942. *J. Bacteriol.* **179**, 5744–5750 (1997).
13. Clark, L. L., Ingall, E. D. & Benner, R. Marine phosphorus is selectively remineralized. *Nature* **393**, 426 (1998).
14. Zehr, J. P. & Ward, B. B. Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl. Environ. Microbiol.* **68**, 1015–1024 (2002).
15. Long, R. A. & Azam, F. Antagonistic interactions among marine pelagic bacteria. *Appl. Environ. Microbiol.* **67**, 4975–4983 (2001).
16. Mann, E. L., Ahlgren, N., Moffett, J. W. & Chisholm, S. W. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol. Oceanogr.* **47**, 976–988 (2002).
17. Palenik, B. & Dyhrman, S. T. In *Phosphorus in Plant Biology: Regulatory Roles in Molecular, Cellular, Organismic, and Ecosystem Processes* (eds Lynch, J. P. & Deikman, J.) 26–38 (American Society of Plant Physiologists, Rockville, MD, 1998).
18. Nyssola, A., Kerovuo, J., Kaukinen, P., von Weymar, N. & Reinikainen, T. Extreme halophiles synthesize betaine from glycine by methylation. *J. Biol. Chem.* **275**, 22196–22201 (2000).
19. Nomura, M., Ishitani, M., Takabe, T., Rai, A. K. & Takabe, T. *Synechococcus* sp. PCC7942 transformed with *Escherichia coli* bet genes produces glycine betaine from choline and acquires resistance to salt stress. *Plant Physiol.* **107**, 703–708 (1995).
20. Badger, M. R., Hanson, D. & Price, G. D. Evolution and diversity of CO₂ concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* **29**, 161–173 (2002).
21. de Lorimier, R., Guglielmi, G., Bryant, D. A. & Stevens, S. E. J. Structure and mutation of a gene encoding a M₃ 33 000 phycocyanin-associated linker polypeptide. *Arch. Microbiol.* **153**, 541–549 (1990).
22. Collier, J. L. & Grossman, A. R. A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J.* **13**, 1039–1047 (1994).
23. Dolganov, N. & Grossman, A. R. A polypeptide with similarity to phycocyanin alpha-subunit phycocyanobilin lyase involved in degradation of phycobilisomes. *J. Bacteriol.* **181**, 610–617 (1999).
24. Volz, K. In *Two-Component Signal Transduction* (eds Hoch, J. A. & Silhavy, T. J.) 53–64 (American Society for Microbiology, Washington DC, 1995).
25. Goto-Seki, A., Shirokane, M., Masuda, S., Tanaka, K. & Takahashi, H. Specificity crosstalk among group 1 and group 2 sigma factors in the cyanobacterium *Synechococcus* sp. PCC7942: *in vitro* specificity and a phylogenetic analysis. *Mol. Microbiol.* **34**, 473–484 (1999).
26. Brahamsha, B. A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Appl. Environ. Microbiol.* **62**, 1747–1751 (1996).
27. Badger, J. H. & Olsen, G. J. CRITICA: Coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* **16**, 512–524 (1999).
28. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**, 4636–4641 (1999).
29. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. *Nucleic Acids Res.* **25**, 955–964 (1997).
30. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the members of the *Prochlorococcus* annotation team: G. Rocap, S. W. Chisholm, D. Lindell, N. Algren, M. Coleman, W. Hess, A. Post, S. Shaw, C. Steglich, C. Ting, M. Sullivan, A. Tolonen, Z. Johnson and E. Zinser. We also thank T. Lane for discussions about carbonic anhydrases. This research was funded by the Biological and Environmental Research Program and the US Department of Energy's Office of Science. Sequencing was carried out and managed at the Joint Genome Institute. Computational analysis was performed at Oak Ridge National Laboratory, managed by UT-BATTELLE for the US Department of Energy. Additional support was provided by a DOE grant to B.P., B.B. and I.P., and an NSF grant to B.B. E.P. and A.D. were supported by the EC program Margenes, and by the Region Bretagne.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.P. (bpalenik@ucsd.edu). The sequence for the chromosome of *Synechococcus* sp. strain WH8102 is deposited in GenBank under accession number BX548020.

Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation

Gabrielle Rocap¹, Frank W. Larimer^{2,3}, Jane Lamerdin³, Stephanie Malfatti³, Patrick Chain^{3,4}, Nathan A. Ahlgren¹, Andrae Arellano³, Maureen Coleman⁵, Loren Hauser^{2,3}, Wolfgang R. Hess^{9*}, Zackary I. Johnson⁵, Miriam Land^{2,3}, Debbie Lindell⁵, Anton F. Post¹⁰, Warren Regala³, Manesh Shah^{2,3}, Stephanie L. Shaw^{6*}, Claudia Steglich⁹, Matthew B. Sullivan⁷, Claire S. Ting⁸, Andrew Tolonen⁷, Eric A. Webb¹¹, Erik R. Zinser⁵ & Sallie W. Chisholm^{5,8}

¹School of Oceanography, University Of Washington, Seattle, Washington 98195, USA

²Computational Biology, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

³Joint Genome Institute, Walnut Creek, California 94598, USA

⁴Lawrence Livermore National Laboratory, Livermore, California 94550, USA

⁵Department of Civil and Environmental Engineering, ⁶Department of Earth, Atmospheric and Planetary Sciences, ⁷Joint Program in Biological Oceanography, Woods Hole Oceanographic Institution, and ⁸Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

⁹Institute of Biology, Humboldt-University, D-10115 Berlin, Germany

¹⁰Interuniversity Institute of Marine Science, 88103 Eilat, Israel

¹¹Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

* Present addresses: Ocean Genome Legacy, Beverly, Massachusetts 01915, USA (W.R.H.); Department of Environmental Science Policy and Management, University of California, Berkeley, California 94720, USA (S.L.S.)

The marine unicellular cyanobacterium *Prochlorococcus* is the smallest-known oxygen-evolving autotroph¹. It numerically dominates the phytoplankton in the tropical and subtropical oceans^{2,3}, and is responsible for a significant fraction of global photosynthesis. Here we compare the genomes of two *Prochlorococcus* strains that span the largest evolutionary distance within the *Prochlorococcus* lineage⁴ and that have different minimum, maximum and optimal light intensities for growth⁵. The high-light-adapted ecotype has the smallest genome (1,657,990 base pairs, 1,716 genes) of any known oxygenic phototroph, whereas the genome of its low-light-adapted counterpart is significantly larger, at 2,410,873 base pairs (2,275 genes). The comparative architectures of these two strains reveal dynamic genomes that are constantly changing in response to myriad selection pressures. Although the two strains have 1,350 genes in common, a significant number are not shared, and these have been differentially retained from the common ancestor, or acquired through duplication or lateral transfer. Some of these genes have obvious roles in determining the relative fitness of the ecotypes in response to key environmental variables, and hence in regulating their distribution and abundance in the oceans.

As an oxyphototroph, *Prochlorococcus* requires only light, CO₂ and inorganic nutrients, thus the opportunities for extensive niche differentiation are not immediately obvious—particularly in view of the high mixing potential in the marine environment (Fig. 1a). Yet co-occurring *Prochlorococcus* cells that differ in their ribosomal DNA sequence by less than 3% have different optimal light intensities for growth⁶, pigment contents⁷, light-harvesting efficiencies⁵, sensitivities to trace metals⁸, nitrogen usage abilities⁹ and cyanophage specificities¹⁰ (Fig. 1b, c). These ‘ecotypes’—distinct genetic lineages with ecologically relevant physiological differences—would be lumped together as a single species on the basis of their rDNA similarity¹¹, yet they have markedly different distributions within a stratified oceanic water column, with high-

Table 1 General features of two *Prochlorococcus* genomes

Genome feature	MED4	MIT9313
Length (bp)	1,657,990	2,410,873
G+C content (%)	30.8	50.7
Protein coding (%)	88	82
Protein coding genes	1,716	2,275
With assigned function	1,134	1,366
Conserved hypothetical	502	709
Hypothetical	80	197
Genes with orthologue in:		
<i>Prochlorococcus</i> MED4	—	1,352
<i>Prochlorococcus</i> MIT9313	1,352	—
<i>Synechococcus</i> WH8102	1,394	1,710
Genes without orthologue in:		
MED4 and WH8102	—	527
MIT9313 and WH8102	284	—
Transfer RNA	37	43
Ribosomal RNA operons	1	2
Other structural RNAs	3	3

light-adapted ecotypes most abundant in surface waters, and their low-light-adapted counterparts dominating deeper waters¹² (Fig. 1a). The detailed comparison between the genomes of two *Prochlorococcus* ecotypes we report here reveals many of the genetic foundations for the observed differences in their physiologies and vertical niche partitioning, and together with the genome of their close relative *Synechococcus*¹³, helps to elucidate the key factors that regulate species diversity, and the resulting biogeochemical cycles, in today's oceans.

The genome of *Prochlorococcus* MED4, a high-light-adapted strain, is 1,657,990 base pairs (bp). This is the smallest of any oxygenic phototroph—significantly smaller than that of the low-

light-adapted strain MIT9313 (2,410,873 bp; Table 1). The genomes of MED4 and MIT9313 consist of a single circular chromosome (Supplementary Fig. 1), and encode 1,716 and 2,275 genes respectively, roughly 65% of which can be assigned a functional category (Supplementary Fig. 2). Both genomes have undergone numerous large and small-scale rearrangements but they retain conservation of local gene order (Fig. 2). Break points between the orthologous gene clusters are commonly flanked by transfer RNAs, suggesting that these genes serve as loci for rearrangements caused by internal homologous recombination or phage integration events.

The strains have 1,352 genes in common, all but 38 of which are also shared with *Synechococcus* WH8102 (ref. 13). Many of the 38 '*Prochlorococcus* -specific' genes encode proteins involved in the atypical light-harvesting complex of *Prochlorococcus*, which contains divinyl chlorophylls *a* and *b* rather than the phycobilisomes that characterize most cyanobacteria. They include genes encoding the chlorophyll *a/b*-binding proteins (*pcb*)¹⁴, a putative chlorophyll *a* oxygenase, which could synthesize (divinyl) chlorophyll *b* from (divinyl) chlorophyll *a*¹⁵, and a lycopene epsilon cyclase involved in the synthesis of alpha carotene¹⁶. This remarkably low number of 'genera defining' genes illustrates how differences in a few gene families can translate into significant niche differentiation among closely related microbes.

MED4 has 364 genes without an orthologue in MIT9313, whereas MIT9313 has 923 that are not present in MED4. These strain-specific genes, which are dispersed throughout the chromosome (Fig. 2), clearly hold clues about the relative fitness of the two strains under different environmental conditions. Almost half of the 923 MIT9313-specific genes are in fact present in *Synechococcus* WH8102, suggesting that they have been lost from MED4 in the course of genome reduction. Lateral transfer events, perhaps

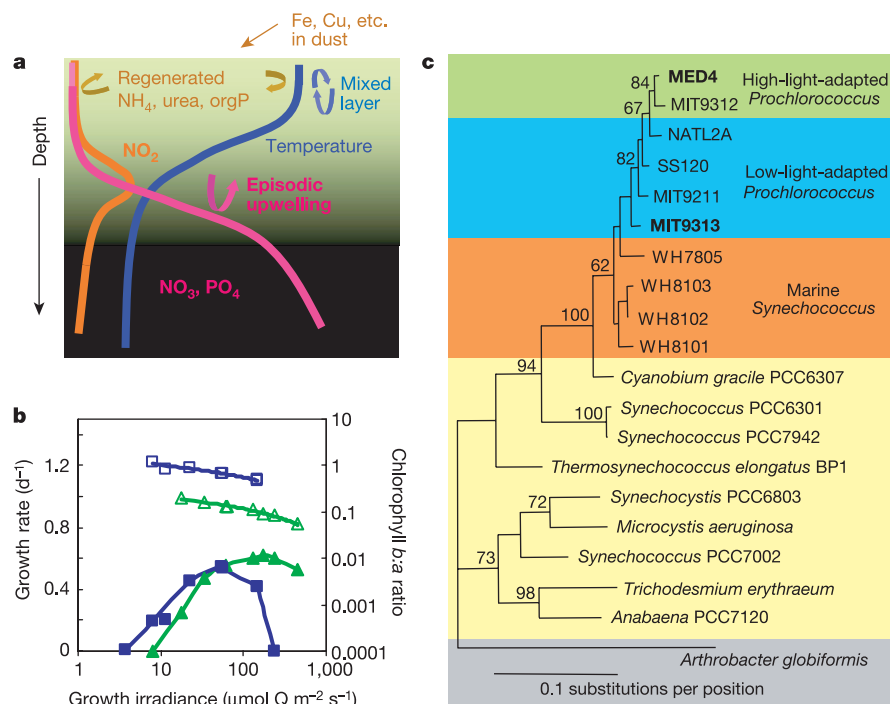


Figure 1 Ecology, physiology and phylogeny of *Prochlorococcus* ecotypes. **a**, Schematic stratified open-ocean water column illustrating vertical gradients allowing niche differentiation. Shading represents degree of light penetration. Temperature and salinity gradients provide a mixing barrier, isolating the low-nutrient/high-light surface layer from the high-nutrient/low-light deep waters. Photosynthesis in surface waters is driven

primarily by rapidly regenerated nutrients, punctuated by episodic upwelling. **b**, Growth rate (filled symbols) and chlorophyll *b*:*a* ratio (open symbols) as a function of growth irradiance for MED4 (ref. 7) (green) and MIT9313 (ref. 6) (blue). **c**, Relationships between *Prochlorococcus* and other cyanobacteria inferred using 16S rDNA.

mediated by phage¹⁰, may also be a source of some of the strain-specific genes (Supplementary Figs 3–6).

Gene loss has played a major role in defining the *Prochlorococcus* photosynthetic apparatus. MED4 and MIT9313 are missing many of the genes encoding phycobilisome structural proteins and enzymes involved in phycobilin biosynthesis¹⁵. Although some of these genes remain, and are functional¹⁷, others seem to be evolving rapidly within the *Prochlorococcus* lineage¹⁸. Selective genome reduction can also be seen in the photosynthetic reaction centre of *Prochlorococcus*. Light acclimation in cyanobacteria often involves differential expression of multiple, but distinct, copies of genes encoding photosystem II D1 and D2 reaction centre proteins (*psbA* and *psbD* respectively)¹⁹. However, MED4 has a single *psbA* gene, MIT9313 has two that encode identical photosystem II D1 polypeptides, and both possess only one *psbD* gene, suggesting a diminished ability to photoacclimate. MED4 has also lost the gene encoding cytochrome *c550* (*psbV*), which has a crucial role in the oxygen-evolving complex in *Synechocystis* PCC6803 (ref. 20).

There are several differences between the genomes that help account for the different light optima of the two strains. For example, the smaller MED4 genome has more than twice as many genes (22 compared with 9) encoding putative high-light-inducible proteins, which seem to have arisen at least in part through duplication events¹⁵. MED4 also possesses a photolyase gene that has been lost in MIT9313, probably because there is little selective pressure to retain ultraviolet damage repair in low light habitats. Regarding differences in light-harvesting efficiencies, it is noteworthy that MED4 contains only a single gene encoding the chlorophyll *a/b*-binding antenna protein *Pcb*, whereas MIT9313 possesses two copies. The second type has been found exclusively in low-light-adapted strains²¹, and may form an antenna capable of binding more chlorophyll pigments.

Both strains have a low proportion of genes involved in regulatory functions. Compared with the freshwater cyanobacterium *Thermosynechococcus elongatus* (genome size <2.6 megabases)²², MIT9313 has fewer sigma factors, transcriptional regulators and two-component sensor-kinase systems, and MED4 is even more reduced (Supplementary Table 1). The circadian clock genes provide an example of this reduction as both genomes lack several components (*pex*, *kaiA*) found in the model *Synechococcus* PCC7942 (ref. 23). However, genes for the core clock proteins (*kaiB*, *kaiC*) remain in both genomes, and *Prochlorococcus* cell division is tightly synchronized to the diel light/dark cycle²⁴. Thus, loss of some circadian components may imply an alternative signalling pathway for circadian control.

Gene loss may also have a role in the lower percentage of G+C content of MED4 (30.8%) compared with that of MIT9313 (50.74%), which is more typical of marine *Synechococcus*. MED4 lacks genes for several DNA repair pathways including recombinational repair (*recJ*, *recQ*) and damage reversal (*mutT*). Particularly, the loss of the base excision repair gene *mutY*, which removes adenosines incorrectly paired with oxidatively damaged guanine residues, may imply an increased rate of G•C to T•A transversions²⁵. The tRNA complement of MED4 is largely identical to MIT9313 and is not optimized for a low percentage G+C genome, suggesting that it is not evolving as fast as codon usage.

Analysis of the nitrogen acquisition capabilities of the two strains points to a sequential decay in the capacity to use nitrate and nitrite during the evolution of the *Prochlorococcus* lineage (Fig. 3a). In *Synechococcus* WH8102—representing the presumed ancestral state—many nitrogen acquisition and assimilation genes are grouped together (Fig. 3a). MIT9313 has lost a 25-gene cluster, which includes genes encoding the nitrate/nitrite transporter and nitrate reductase. The nitrite reductase gene has been retained in MIT9313, but it is flanked by a proteobacterial-like nitrite transporter rather than a typical cyanobacterial nitrate/nitrite permease (Supplementary Fig. 4), suggesting acquisition by lateral gene

transfer. An additional deletion event occurred in MED4, in which the nitrite reductase gene was also lost (Fig. 3a). As a result of these serial deletion events MIT9313 cannot use nitrate, and MED4 cannot use nitrate or nitrite⁹. Thus each *Prochlorococcus* ecotype uses the N species that is most prevalent at the light levels to which they are best adapted: ammonium in the surface waters and nitrite at depth (Fig. 1a). *Synechococcus*, which is the only one of the three that has nitrate reductase, is able to bloom when nitrate is upwelled (Fig. 1a), as occurs in the spring in the North Atlantic³ and the north Red Sea²⁶.

The two *Prochlorococcus* strains are also less versatile in their organic N usage capabilities than *Synechococcus* WH8102 (ref. 13). MED4 contains the genes necessary for usage of urea, cyanate and oligopeptides, but no monomeric amino acid transporters have been identified. In contrast, MIT9313 contains transporters for urea, amino acids and oligopeptides but lacks the genes necessary for cyanate usage (cyanate transporter and cyanate lyase) (Fig. 3a). As expected, both genomes contain the high-affinity ammonium transporter *amt1* and both lack the nitrogenase genes essential for nitrogen fixation. Finally, both contain the nitrogen transcriptional regulator encoded by *ntcA* and there are numerous genes in both genomes, including *ntcA*, *amt1*, the urea transport and GS/GOGAT genes (glutamine synthetase and glutamate synthase, both involved in ammonia assimilation), with an upstream *NtcA*-binding-site consensus sequence.

The genomes also have differences in genes involved in phosphorus usage that have obvious ecological implications. MED4, but not MIT9313, is capable of growth on organic P sources (L. R. Moore and S.W.C., unpublished data), and organic P can be the prevalent form of P in high-light surface waters²⁷. This difference may be due to the acquisition of an alkaline phosphatase-like gene in MED4 (Supplementary Fig. 5). Both genomes contain the high-affinity phosphate transport system encoded by *pstS* and *pstABC*²⁸, but MIT9313 contains an additional copy of the phosphate-binding component *pstS*, perhaps reflecting an increased reliance on orthophosphate in deeper waters. MED4 contains

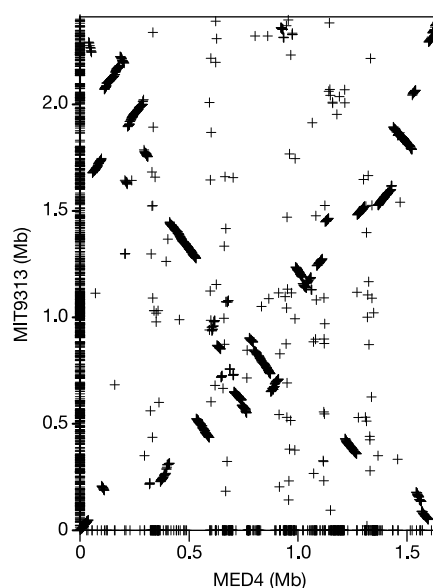
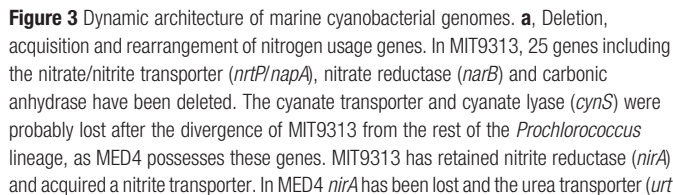


Figure 2 Global genome alignment as seen from start positions of orthologous genes. Genes present in one genome but not the other are shown on the axes. The 'broken X' pattern has been noted before for closely related bacterial genomes, and is probably due to multiple inversions centred around the origin of replication. Alternating slopes of many adjacent gene clusters indicate that multiple smaller-scale inversions have also occurred.

Prochlorococcus does not use typical cyanobacterial genes for inorganic carbon concentration or fixation. Both genomes contain a sodium/bicarbonate symporter but lack homologues to known

Prochlorococcus has been identified in deep suboxic zones where it is unlikely that they can sustain themselves by photosynthesis alone²⁹, thus we looked for genomic evidence of heterotrophic capability. Indeed, the presence of oligopeptide transporters in both genomes, and the larger proportion of transporters (including some sugar transporters) in the MIT9313 strain-specific genes (Supplementary Fig. 2), suggests the potential for partial hetero-



cluster) and urease (*ure* cluster) genes have been rearranged (dotted line). Genes in different functional categories are colour-coded to guide the eye. **b**, Lateral transfer of genes involved in lipopolysaccharide biosynthesis including sugar transferases, sugar epimerases, modifying enzymes and two pairs of ABC-type transporters. Blue, genes in all three genomes; pink, genes hypothesized to have been laterally transferred; red, tRNAs; white, other genes. The percentage of G + C content in MIT9313 along this segment is lower (42%) than the whole-genome average (horizontal line).

trophy. However, neither genome contains known pathways that would allow for complete heterotrophy. They are both missing genes for steps in the tricarboxylic acid cycle, including 2-oxoglutarate dehydrogenase, succinyl-CoA synthetase and succinyl-CoA-acetoacetate-CoA transferase.

Cell surface chemistry has a major role in phage recognition and grazing by protists and thus is probably under intense selective pressure in nature. The two *Prochlorococcus* genomes and the *Synechococcus* WH8102 genome show evidence of extensive lateral gene transfer and deletion events of genes involved in lipopolysaccharide and/or surface polysaccharide biosynthesis, reinforcing the role of predation pressures in the creation and maintenance of microdiversity. For example, MIT9313 has a 41.8-kilobase (kb) cluster of surface polysaccharide genes (Fig. 3b), which has a lower percentage G+C composition (42%) than the genome as a whole, implicating acquisition by lateral gene transfer. MED4 has acquired a 74.5-kb cluster consisting of 67 potential surface polysaccharide genes (Supplementary Fig. 6a) and has lost another cluster of surface polysaccharide biosynthesis genes shared between MIT9313 and *Synechococcus* WH8102 (Supplementary Fig. 6b).

The approach we have taken in describing these genomes highlights the known drivers of niche partitioning of these closely related organisms (Fig. 1). Detailed comparisons with the genomes of additional strains, such as *Prochlorococcus* SS120 (ref. 30), will enrich this story, and the analysis of whole genomes from *in situ* populations will be necessary to understand the full expanse of genomic diversity in this group. The genes of unknown function in all of these genomes hold important clues for undiscovered niche dimensions in the marine pelagic zone. As we unveil their function we will undoubtedly learn that the suite of selective pressures that shape these communities is much larger than we have imagined. Finally, it may be useful to view *Prochlorococcus* and *Synechococcus* as important 'minimal life units', as the information in their roughly 2,000 genes is sufficient to create globally abundant biomass from solar energy and inorganic compounds. □

Methods

Genome sequencing and assembly

DNA was isolated from the clonal, axenic strain MED4 and the clonal strain MIT9313 essentially as described previously⁴. The two whole-genome shotgun libraries were obtained by fragmenting genomic DNA using mechanical shearing and cloning 2–3-kb fragments into pUC18. Double-ended plasmid sequencing reactions were carried out using PE BigDye Terminator chemistry (Perkin Elmer) and sequencing ladders were resolved on PE 377 Automated DNA Sequencers (Perkin Elmer). The whole-genome sequence of *Prochlorococcus* MED4 was obtained from 27,065 end sequences (7.3-fold redundancy), whereas *Prochlorococcus* MIT9313 was sequenced to X6.2 coverage (33,383 end sequences). For *Prochlorococcus* MIT9313, supplemental sequencing (X0.05 sequence coverage) of a pFos1 fosmid library was used as a scaffold. Sequence assembly was accomplished using PHRAP (P. Green). All gaps were closed by primer walking on gap-spanning library clones or PCR products. The final assembly of *Prochlorococcus* MED4 was verified by long-range genomic PCR reactions, whereas the assembly of *Prochlorococcus* MIT9313 was confirmed by comparison to the fosmid clones, which were fingerprinted with EcoRI. No plasmids were detected in the course of genome sequencing, and insertion sequences, repeated elements, transposons and prophages are notably absent from both genomes. The likely origin of replication in each genome was identified based on G+C skew, and base pair 1 was designated adjacent to the *dnaN* gene.

Genome annotation

The combination of three gene-modelling programs, Critica, Glimmer and Generation, were used in the determination of potential open reading frames and were checked manually. A revised gene/protein set was searched against the KEGG GENES, Pfam, PROSITE, PRINTS, ProDom, COGs and CyanoBase databases, in addition to BLASTP against the non-redundant peptide sequence database from GenBank. From these results, categorizations were developed using the KEGG and COGs hierarchies, as modified in CyanoBase. Manual annotation of open reading frames was done in conjunction with the *Synechococcus* team. The three-way genome comparison was used to refine predicted start sites, add additional open reading frames and standardize the annotation across the three genomes.

Genome comparisons

The comparative genome architecture of MED4 and MIT9313 was visualized using the Artemis Comparison Tool (<http://www.sanger.ac.uk/Software/ACT/>). Orthologues were determined by aligning the predicted coding sequences of each gene with the coding

sequences of the other genome using BLASTP. Genes were considered orthologues if each was the best hit of the other one and both *e*-values were less than e^{-10} . In addition, bidirectional best hits with *e*-values less than e^{-6} and small proteins of conserved function were manually examined and added to the orthologue lists.

Phylogenetic analyses used PAUP*, logdet distances and minimum evolution as the objective function. The degree of support at each node was evaluated using 1,000 bootstrap resamplings. Ribosomal DNA analyses used 1,160 positions. The Gram-positive bacterium *Arthrobacter globiformis* was used to root the tree.

Received 11 May; accepted 25 July 2003; doi:10.1038/nature01947.

Published online 13 August 2003.

- Chisholm, S. W. *et al.* A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**, 340–343 (1988).
- Campbell, L., Liu, H., Nolla, H. & Vaulot, D. Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991–1994 ENSO event. *Deep Sea Res. II* **44**, 167–192 (1997).
- DuRand, M. D., Olson, R. J. & Chisholm, S. W. Phytoplankton population dynamics at the Bermuda Atlantic Time-series Station in the Sargasso Sea. *Deep Sea Res. II* **48**, 1983–2003 (2001).
- Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S rDNA internal transcribed spacer (ITS) sequences. *Appl. Environ. Microbiol.* **68**, 1180–1191 (2002).
- Moore, L. R. & Chisholm, S. W. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr.* **44**, 628–638 (1999).
- Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464–467 (1998).
- Moore, L. R., Goericke, R. E. & Chisholm, S. W. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar. Ecol. Prog. Ser.* **116**, 259–275 (1995).
- Mann, E. L., Ahlgren, N., Moffett, J. W. & Chisholm, S. W. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol. Oceanogr.* **47**, 976–988 (2002).
- Moore, L. R., Post, A. F., Rocap, G. & Chisholm, S. W. Utilization of different nitrogen sources by the marine cyanobacteria, *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **47**, 989–996 (2002).
- Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
- Hagström, Å. *et al.* Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl. Environ. Microbiol.* **68**, 3628–3633 (2002).
- West, N. J. *et al.* Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by *in situ* hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* **147**, 1731–1744 (2001).
- Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424**, 1037–1042 (2003).
- La Roche, J. *et al.* Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light harvesting proteins. *Proc. Natl Acad. Sci. USA* **93**, 15244–15248 (1996).
- Hess, W. *et al.* The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics. *Photosynth. Res.* **70**, 53–71 (2001).
- Stickforth, P., Steiger, S., Hess, W. R. & Sandmann, G. A novel type of lycopene ϵ -cyclase in the marine cyanobacterium *Prochlorococcus marinus* MED4. *Arch. Microbiol.* **179**, 407–415 (2003).
- Frankenberg, N., Mukougawa, K., Kohchi, T. & Lagarias, J. C. Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *Plant Cell* **13**, 965–978 (2001).
- Ting, C., Rocap, G., King, J. & Chisholm, S. W. Phycobiliprotein genes of the marine prokaryote *Prochlorococcus*: Evidence for rapid evolution of genetic heterogeneity. *Microbiology* **147**, 3171–3182 (2001).
- Golden, S. S., Brusslan, J. & Haselkorn, R. Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. *EMBO J.* **5**, 2789–2798 (1986).
- Shen, J. R., Qian, M., Inoue, Y. L. & Burnap, R. L. Functional characterization of *Synechocystis* sp. PCC 6803 $\Delta psbU$ and $\Delta psbV$ mutants reveals important roles of cytochrome *c*-550 in cyanobacterial oxygen evolution. *Biochemistry* **37**, 1551–1558 (1998).
- Garczarek, L., van der Staay, G. W. M., Hess, W. R., Le Gall, F. & Partensky, F. Expression and phylogeny of the multiple antenna genes of the low-light-adapted strain *Prochlorococcus marinus* SS120 (Oxyphotobacteria). *Plant Mol. Biol.* **46**, 683–693 (2001).
- Nakamura, Y. *et al.* Complete genome structure of the thermophilic Cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res.* **9**, 123–130 (2002).
- Ishiura, M. *et al.* Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* **281**, 1519–1523 (1998).
- Vaulot, D., Marie, D., Olson, R. J. & Chisholm, S. W. Growth of *Prochlorococcus*, a photosynthetic prokaryote, in the Equatorial Pacific Ocean. *Science* **268**, 1480–1482 (1995).
- Michaels, M. L. & Miller, J. H. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-Hydroxyguanine (7,8-Dihydro-8-Oxoguanine). *J. Bacteriol.* **174**, 6321–6325 (1992).
- Lindell, D. & Post, A. F. Ultraphytoplankton succession is triggered by deep winter mixing in the Gulf of Aqaba (Eilat). *Red Sea. Limnol. Oceanogr.* **40**, 1130–1141 (1995).
- Karl, D. M., Bidigare, R. R. & Letelier, R. M. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: The domain shift hypothesis. *Deep Sea Res. II* **48**, 1449–1470 (2001).
- Scanlan, D. J., Mann, N. H. & Carr, N. G. The response of the picoplanktonic marine cyanobacterium *Synechococcus* species WH7803 to phosphate starvation involves a protein homologous to the periplasmic phosphate binding protein of *Escherichia coli*. *Mol. Microbiol.* **10**, 181–191 (1993).
- Johnson, Z. *et al.* Energetics and growth kinetics of a deep *Prochlorococcus* spp. population in the Arabian Sea. *Deep Sea Res. II* **46**, 1719–1743 (1999).
- Dufresne, A. *et al.* Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a near minimal oxyphototrophic genome. *Proc. Natl Acad. Sci. USA* (in the press).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This research was funded by the Biological and Environmental Research Program of the US Department of Energy's Office of Science. The Joint Genome Institute managed the overall sequencing effort. Genome finishing was carried out under the auspices of the US Department of Energy by the University of California, Lawrence Livermore National Laboratory. Computational annotation was carried out at the Oak Ridge National Laboratory, managed by UT-BATTELLE for the US Department of Energy. Additional support was provided by the DOE, NSF and the Seaver Foundation to S.W.C., the Israel-US Binational Science Foundation to A.F.P. and S.W.C., and FP5-Margenes to W.R.H. and A.F.P. We thank the *Synechococcus* WH8102 annotators (B. Palenik, B. Brahamsha, J. McCarren, E. Allen, F. Partensky, A. Dufresne and I. Paulsen) for their help with curating the *Prochlorococcus* genomes and E. V. Armbrust and L. Moore for critical reading of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.W.C. (chisholm@mit.edu). The complete nucleotide sequences and sequences of predicted open reading frames have been deposited in the EMBL/GenBank/DBJ databases under accession numbers BX548174 (MED4) and BX548175 (MIT9313).

Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*

Matthew B. Sullivan¹, John B. Waterbury³ & Sallie W. Chisholm²

¹MIT/Woods Hole Oceanographic Institution Joint Program in Biological Oceanography, and ²Department of Civil and Environmental Engineering and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

³Woods Hole Oceanographic Institution, Department of Biology, Woods Hole, Massachusetts 02543, USA

Prochlorococcus is the numerically dominant phototroph in the tropical and subtropical oceans, accounting for half of the photosynthetic biomass in some areas^{1,2}. Here we report the isolation of cyanophages that infect *Prochlorococcus*, and show that although some are host-strain-specific, others cross-infect with closely related marine *Synechococcus* as well as between high-light- and low-light-adapted *Prochlorococcus* isolates, suggesting a mechanism for horizontal gene transfer. High-light-adapted *Prochlorococcus* hosts yielded *Podoviridae* exclusively, which were extremely host-specific, whereas low-light-adapted *Prochlorococcus* and all strains of *Synechococcus* yielded primarily *Myoviridae*, which has a broad host range. Finally, both *Prochlorococcus* and *Synechococcus* strain-specific cyanophage titres were low (<10⁵ ml⁻¹) in stratified oligotrophic waters even where total cyanobacterial abundances were high (>10⁵ cells ml⁻¹). These low titres in areas of high total host cell abundance seem to be a feature of open ocean ecosystems. We hypothesize that gradients in cyanobacterial population diversity, growth rates, and/or the incidence of lysogeny underlie these trends.

Phages are thought to evolve by the exchange of genes drawn from a common gene pool through differential access imposed by host range limitations³. Similarly, horizontal gene transfer, important in microbial evolution^{4,5}, can be mediated by phages⁶ and is probably responsible for many of the differences in the genomes of closely related microbes⁷. Recent detailed analyses of molecular phylogenies constructed for marine *Prochlorococcus* and *Synechococcus*^{7,8} (Fig. 1) show that these genera form a single group within the marine picophytoplankton clade⁹ (>96% identity in 16S ribosomal DNA sequences), yet display microdiversity in the form of ten well-defined subgroups⁸. We have used members of these two groups to study whether phage isolated on a particular host strain cross-infect other hosts, and if so, whether the probability of cross-infection is related to rDNA-based evolutionary distance between the hosts.

Analyses of host range were conducted (Fig. 1) with 44 cyanophages, isolated as previously described¹⁰ from a variety of water depths and locations (see Supplementary Information) using 20 different host strains chosen to represent the genetic diversity of *Prochlorococcus* and *Synechococcus*⁸. Although we did not examine how these patterns would change if phage were propagated on different hosts, this would undoubtedly add another layer of complexity due to host range modifications as a result of methylation of phage DNA⁶. Similar to those that infect other marine bacteria¹¹ and *Synechococcus*¹⁰⁻¹⁴, our *Prochlorococcus* cyanophage isolates fell into three morphological families: *Myoviridae*, *Siphoviridae* and *Podoviridae*¹⁵.

As would be predicted¹⁰⁻¹⁴, *Podoviridae* were extremely host specific with only two cross-infections out of a possible 300 (Fig. 1). Similarly, the two *Siphoviridae* isolated were specific to their hosts. In instances of extreme host specificity, *in situ* host abundance would need to be high enough to facilitate phage-host contact. It is noteworthy in this regard that members of the high-light-adapted *Prochlorococcus* cluster, which yielded the most host-specific cyanophage, have high relative abundances *in situ*¹⁶. The *Myoviridae* exhibited much broader host ranges, with 102 cross-infections out of a possible 539. They not only cross-infected among and between *Prochlorococcus* ecotypes but also between *Prochlorococcus* and *Synechococcus*. Those isolated with *Synechococcus* host strains have broader host ranges and are more likely to cross-infect low-light-adapted than high-light-adapted *Prochlorococcus* strains. The low-light-adapted *Prochlorococcus* are less diverged from *Synechococcus* than high-light-adapted *Prochlorococcus*^{7,8}, suggesting a relationship, in this instance, between the probability of cross-infection and rDNA relatedness of hosts. Finally, we tested the *Myoviridae* for cross-infection against marine bacterial isolates closely related to *Pseudoalteromonas*, which are known to be broadly susceptible to diverse bacteriophages (bacterial strains HER1320, HER1321, HER1327, HER1328)¹¹. None of the *Myoviridae* cyanophages infected these bacteria.

Phage morphotypes isolated were determined, to some degree, by the host used for isolation (Fig. 1). For example, ten of ten cyanophages isolated using high-light-adapted *Prochlorococcus* strains were *Podoviridae*. In contrast, all but two cyanophages isolated on *Synechococcus* were *Myoviridae*, a bias that has been reported by others¹⁴, and over half of those isolated on low-light-adapted *Prochlorococcus* belonged to this morphotype. We further substantiated these trends by examining lysates (as opposed to plaque-purified isolates) from a range of host strains, geographic locations and depths—of 58 *Synechococcus* lysates 93% contained *Myoviridae*, of 43 low-light-adapted *Prochlorococcus* lysates 65% contained *Myoviridae*, and of 107 high-light-adapted *Prochlorococcus* lysates 98% contained *Podoviridae* (see Supplementary Information).

Maximum cyanophage titres, using a variety of *Synechococcus* hosts, are usually found to be within an order of magnitude of the total *Synechococcus* abundance^{10,14,17,18}, and can be as high as 10⁶ phage ml⁻¹. One study¹⁷ has shown, for example, that along a transect in which total *Synechococcus* abundance decreased from 10⁵ cells ml⁻¹ to 250 cells ml⁻¹, maximum cyanophage titres remained at least as high as the total number of *Synechococcus*. We wondered whether titres of *Prochlorococcus* cyanophage in the Sargasso Sea, where *Prochlorococcus* cells are abundant (10⁵ cells ml⁻¹), would be comparable to those measured in coastal oceans for *Synechococcus* where total *Synechococcus* host abundances are of similar magnitude. We assayed cyanophage titres in a depth profile in the Sargasso Sea at the end of seasonal stratification using 11 strains of *Prochlorococcus* (Fig. 2), choosing at least one host strain from each of the six phylogenetic clusters that span the rDNA-based genetic diversity of our culture collection⁸.

Three *Prochlorococcus* host strains (MIT 9303, MIT 9313 and SS120) yielded low or no cyanophage. Other hosts yielded titres

that reached a maximum at 70 m (NATL2A-phage) or 100 m (MIT 9302-, MIT 9515-, MED4-, MIT 9211-, NATL1A-phage) near the depth of maximum *Prochlorococcus* abundance (Fig. 2). All *Prochlorococcus* cyanophage titres were low (<350 cyanophage ml^{-1}) compared with those reported for *Synechococcus* in coastal regions (approximately 10^4 – 10^6 cyanophage ml^{-1}) even though total host abundances were similar between these regions (approximately 10^5 cells ml^{-1})^{10,14,17,18}. *Prochlorococcus* cyanophage titres are comparable to those of *Synechococcus* from oligotrophic waters in the Gulf of Mexico—but in that instance the total *Synechococcus* abundance was also low (<250 cells ml^{-1})¹⁷.

Cyanophage titres were also examined along a surface water transect from coastal (mesotrophic) to open ocean (oligotrophic) in the Atlantic Ocean to better understand the relationship between maximum phage titre and total host abundance along a trophic gradient. Titres were assayed with 12 strains of *Synechococcus* and *Prochlorococcus* that represented the known rDNA-based genetic diversity at the time that we began the study⁸ (but see also ref. 19). We found that *Synechococcus* cyanophage titres decreased by an order of magnitude or greater in surface waters between the coastal and open ocean (Sargasso) sites, whereas total *Synechococcus* abundance decreased from 3×10^4 to 7×10^3 cells ml^{-1} (Fig. 3). *Prochlorococcus* hosts did not yield cyanophage in coastal samples where there are no *Prochlorococcus* cells, and yielded relatively low

titres (0 to 1.5×10^3 phage ml^{-1}) at the shelf, slope and Sargasso stations where total *Prochlorococcus* abundance was between 4.5×10^4 and 1.4×10^5 cells ml^{-1} . Even though total *Prochlorococcus* abundance at the Sargasso site was similar to that of *Synechococcus* at the coastal site (Fig. 3i, j), *Prochlorococcus* and *Synechococcus* cyanophage titres were significantly lower at the open ocean site (Fig. 3a–h). Moreover, regardless of the host used, titres never exceeded 3×10^3 cyanophage ml^{-1} at any depth throughout the photic zone even though total *Prochlorococcus* abundances exceeded 10^5 cells ml^{-1} (see Supplementary Information). Thus it seems that cyanophage titres at the end of summer stratification are relatively low in open ocean ecosystems, where the total possible host cell abundances are relatively high. Low titres lead to reduced contact rates and lowered mortality rates⁶.

Although it is difficult to draw definitive conclusions about causality from such trends because of the complexity of the phage–host interaction, there are some factors that might be implicated. If, for example, host strain microdiversity increased along the transect and cross-infection ability did not increase concurrently, this would lead to lower phage titres yielded by a suite of host strains²⁰. Indeed, we know that the relative abundance of *Synechococcus* ecotypes changes from coastal to oligotrophic waters¹⁶. However, we observed a systematic decrease in cyanophage titres for all five *Synechococcus* hosts (note WH 8020 yielded no

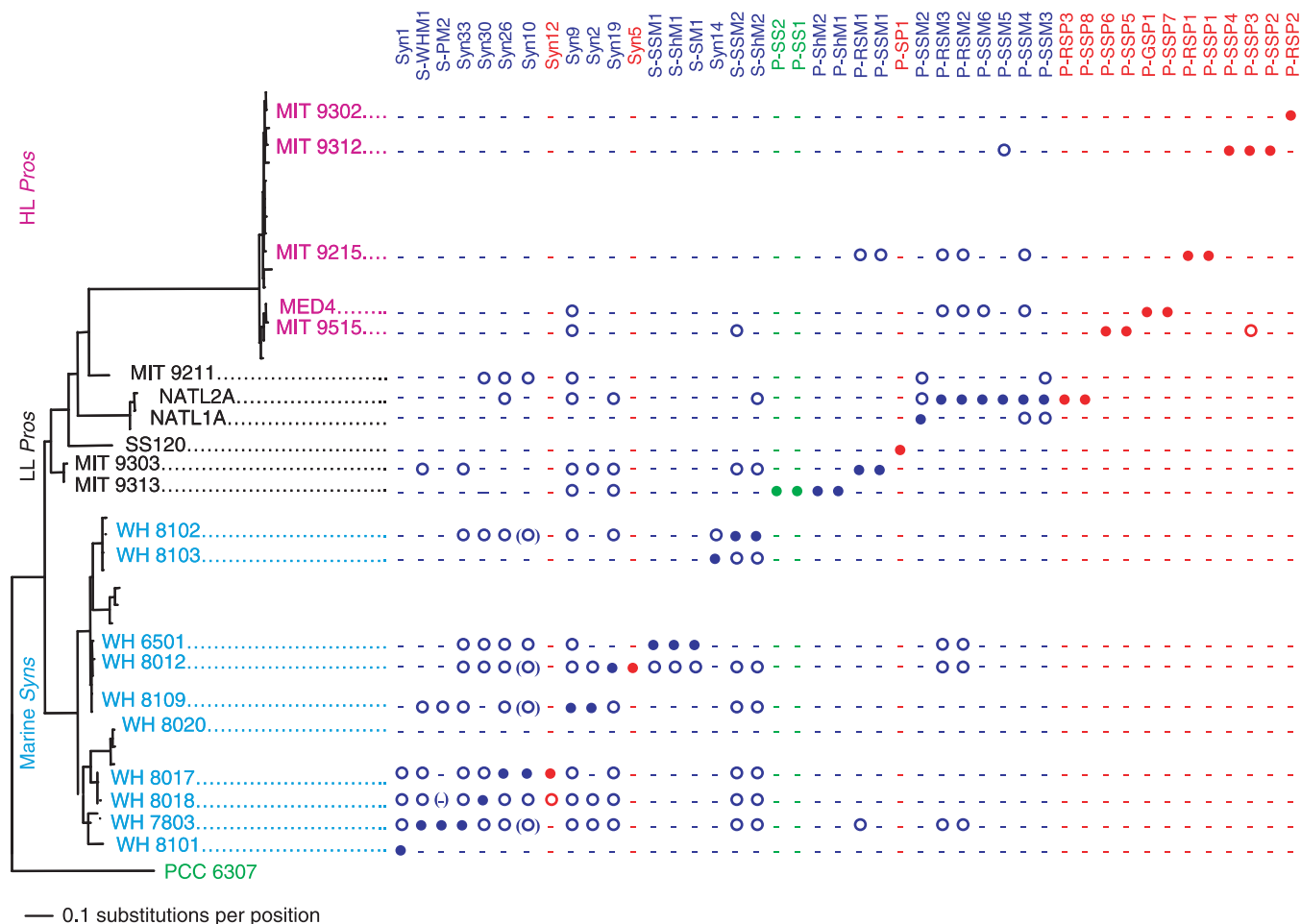


Figure 1 Host ranges of 44 clonal cyanophages exposed to marine *Prochlorococcus* and *Synechococcus* cultured isolates. The evolutionary relationships between the 21 host strains are shown in the phylogenetic tree inferred using 16S–23S rDNA spacer regions⁸. For the cyanophages, red indicates *Podoviridae*, blue indicates *Myoviridae* and green indicates *Siphoviridae*. Filled circles, host strain used to isolate a particular cyanophage;

open circles, cross-infection of cyanophage with another host; dash, no infection (that is, lysis). Symbols in parentheses indicate that the results do not match earlier studies^{10,13} with these phage and hosts (see Methods for details). HL Pros, high-light-adapted *Prochlorococcus*; LL Pros, low-light-adapted *Prochlorococcus*; Marine Syms, marine *Synechococcus*.

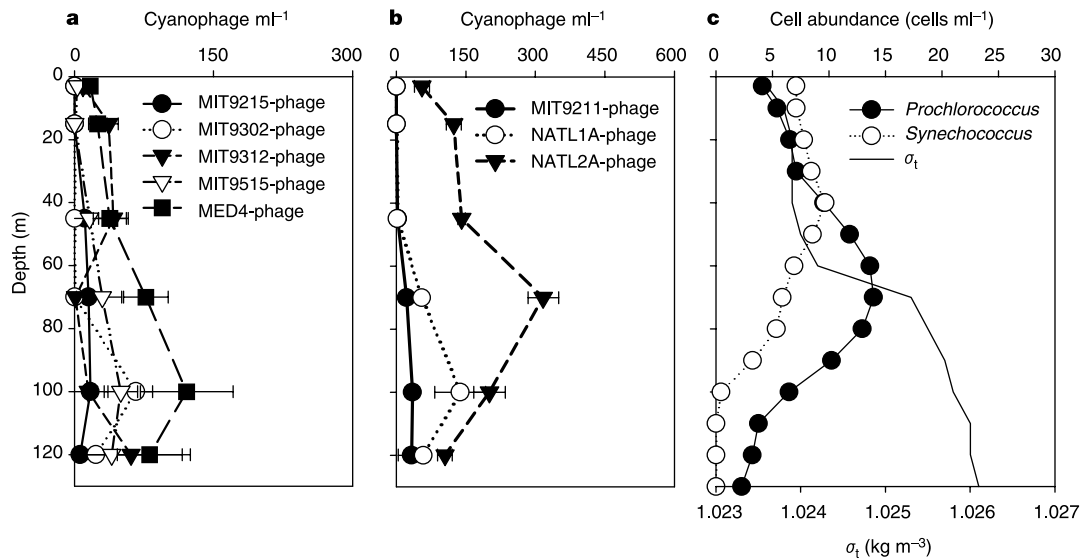


Figure 2 Cyanophage titres, measured using *Prochlorococcus* host strains, as a function of depth at the Bermuda Atlantic Time Series Station in the Sargasso Sea on 26 September 1999. Nine of the cyanophages used in host range analyses were isolated from this depth profile (see Supplementary Information). **a–c**, Titres measured using high-light-adapted *Prochlorococcus* hosts (**a**) and low-light-adapted hosts (**b**), and total

Prochlorococcus and *Synechococcus* cell abundances (*Prochlorococcus* cells $\times 10^4$ ml⁻¹; *Synechococcus* cells $\times 10^3$ ml⁻¹) and σ_t (a proxy for water density used to measure the depth of the mixed layer) (**c**). Titres were undetectable for low-light-adapted *Prochlorococcus* strains SS120, MIT9303 and MIT9313. Error bars represent the s.d. of assays.

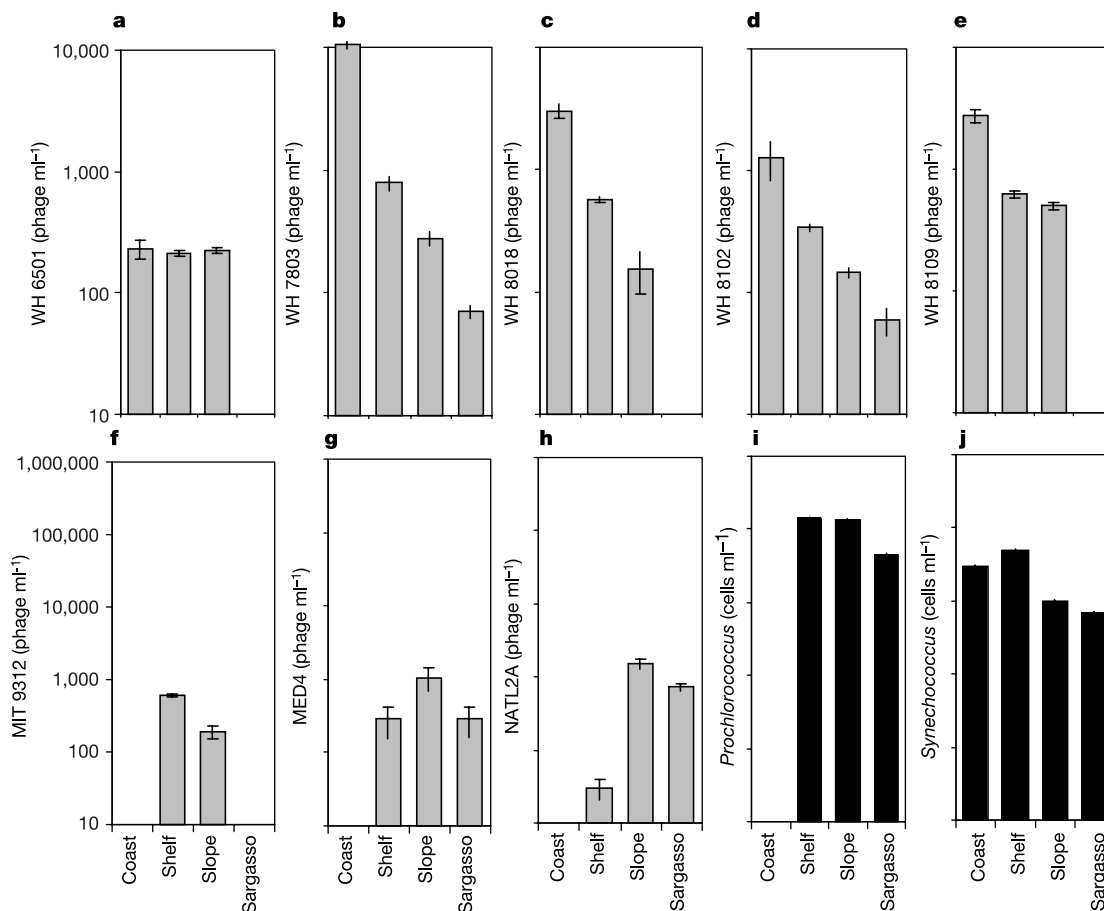


Figure 3 Cyanophage titres measured in *Synechococcus* and *Prochlorococcus* host cells along a surface water transect from coastal (coast, Woods Hole, Massachusetts) to open ocean (Sargasso) conducted in September 2001. Note that the magnitudes of the y-axes are different for **a–e** and **f–j**. Ten of the cyanophages used in host range analyses were isolated from along this transect (see Supplementary Information). **a–h**, Cyanophage

titres represent the averages and s.d. of triplicate plaque assays. **i, j**, Cell concentrations represent averages and s.d. of duplicate flow cytometry assays. Where no bar is shown, there were no plaques (**a, c, e–h**) or no cells (**i**). No plaques were observed at any of the surface samples along the transect for *Synechococcus* strain WH 8020 and *Prochlorococcus* strains MIT 9313, SS120 and MIT 9211 (data not shown).

plaques) at the extremes of this transect (the coastal and Sargasso sites; Fig. 3a–e). If changing host abundance alone explained the change in titres, and if our suite of host strains is representative of natural diversity, then one might expect at least one host strain to yield increasing titres along the gradient (for example, for the ‘open ocean strain’ WH 8102); however, this was not observed.

Another possible explanation for decreasing phage titres as one goes from coastal to open ocean ecosystems is decreased nutrient availability along the transect²¹ resulting in suboptimal growth of host cells in the Sargasso Sea²² relative to *Synechococcus* at the coastal site²³. Viral production is correlated with host growth rates in chemostats²⁴ and in the field²⁵, which could result from nutrient limitation causing physiological changes in the host that stall the lytic process of obligately lytic phage⁶, or favour lysogeny in temperate phage^{18,26,27}. Although temperate phage have not been identified for marine *Prochlorococcus* or *Synechococcus*, INT family site-specific recombinases exist in the genomes of *Prochlorococcus* MED4 and MIT 9313, and *Synechococcus* WH 8102 (http://www.jgi.doe.gov/JGI_microbial/html/index.html), suggesting that prophages were once integrated into these host genomes^{28,29}.

The phage–host system described here should continue to be a useful framework for advancing our understanding of the ecology and evolution of phage–host interactions in marine ecosystems. We have known for some time that cyanophages must have a role in maintaining genetic diversity among hosts^{10,17}. The broad host ranges reported here indicate further their potential for mediating horizontal gene transfer, which may help explain the extensive microdiversity^{8,19,28,29} seen in these two groups of marine cyanobacteria. The extent to which this potential is realized should become clear as more and more host and phage genomes are sequenced. Of significance also is the coupling between phage morphology and host type. Experiments designed to characterize phage resistance across variable hosts and phages (for example, identification of receptors and restriction and modification systems) should elucidate the underlying mechanisms responsible for these patterns. Finally, our analyses of cyanophage titres along a coastal–open ocean transect suggest that the underlying processes responsible for the production of free cyanophage differ along trophic gradients in the oceans. To fully explain these observations will require the development of approaches that allow one to determine which phage can infect which host(s) in a given community, and an understanding of the relative roles of lytic and lysogenic phases of the viral life cycle in aquatic systems. □

Methods

Sample collection

Water samples for cyanophage titres, cyanophage isolations and cyanobacterial abundances were collected at the Bermuda Atlantic Time Series Station on 26 September 1999 and at four sites along a transect from Woods Hole to the Western Sargasso Sea on 5, 16, 17 and 22 September 2001 (see Supplementary Information). Water for cyanophage isolations was filtered (0.4 µm, Poretics number 13028 in 1999; 0.2 µm, Osmonics number K02CP04700 in 2001) and stored at 4 °C in the dark in acid-washed polycarbonate (1999) or glass (2001) bottles until analysis (up to 15 months later). Cyanophage titres remain stable for at least one year²⁷. Control experiments showed that titres were stable over a 15-month period (see Supplementary Information).

Culturing conditions

Prochlorococcus and *Synechococcus* strains were maintained in ‘75% Pro99’ medium, a modification of the ‘Pro2’ medium³⁰ with a 75% seawater base and the following final concentrations of N and P: 800 µM NH₄Cl, 50 µM NaH₂PO₄. Cultures were grown at 19–21 °C under constant light 8–12 µE m^{−2} s^{−1} for low-light-adapted *Prochlorococcus* and *Synechococcus*; 35–45 µE m^{−2} s^{−1} for high-light-adapted *Prochlorococcus*.

Cyanophage isolations

Prochlorococcus cyanophage isolations were done initially using an axenic strain of *Prochlorococcus* (MED4ax; M. Saito and J.B.W., unpublished observations). Exponentially growing cells were transferred to fresh medium (1 ml:20 ml) and inoculated with 1 ml of 0.4-µm-filtered sea water. The time course of auto-fluorescence (chlorophyll biomass) of these cultures was then followed with a Turner Designs 10-AU fluorometer. Cultures showing reduced fluorescence relative to controls were filtered and examined for phage particles as previously described^{10,12}. Lysates were stored at 4 °C in the dark. Subsequent

isolations using 19 additional host strains were done using the same procedures scaled down to small volumes. Cyanophage isolates used in this study were plaque-purified twice before use, classified using morphology described by the ICTV¹⁵, and named according to suggestions made for cyanophage¹⁸.

Cyanophage host range

Host range analyses were conducted over a period of about 2 yr. Each interaction between a cyanophage and its potential host cell was performed with exponentially growing cells in triplicate on at least two different occasions. Marine bacterial strains were purchased from the Felix d’Herelle Reference Center for Bacterial Viruses (contact H. Ackermann). Several of the *Synechococcus* cyanophage used in this study (‘Syn’ phages; S-PM2 and S-WHM1) had been previously examined for host range cross-infectivity^{10,13} and were maintained as lysates at 4 °C in the dark while host cyanobacterial cultures were serially transferred in late exponential and early stationary phase. A total of 103 of 108 cross-infections using these stored cyanophages yielded similar host range results in this study (Fig. 1). Of the differences observed, four of five were for one cyanophage isolate (Syn10), suggesting that it might have evolved an extended host range mutation. Host range can be altered through DNA modifications that can occur during propagation of a phage on an alternative host. Overall, these results suggest that cyanophage susceptibility of these host strains and the cross-infectivity of the cyanophage remained relatively stable throughout the 10 or more years of storage and culture maintenance.

Host cell and cyanophage quantification

Host cell abundance was measured using a modified Becton-Dickinson FACSscan flow cytometer⁷. Cyanophage titres were quantified using most probable number (MPN) assays (1999) or plaque assay (2001). MPN assays were monitored for lysis relative to controls for 2–3 weeks depending on the host strain used. For the plaque assays, we plated the host strain in soft agarose (0.4% final concentration; GIBCO BRL, Life Technologies number 5517-014) along with the phage being titred; lawns of cells appeared 8–28 days after inoculation, depending on the strain (M. Saito, M.B.S. and J.B.W., unpublished data). Plaques were counted daily until they no longer appeared (3–14 days after the first plaques). Titres measured using both assays were not significantly different (*t*-test assuming equal variances, $\alpha = 0.10$).

Host dependency of measured titres

To see whether our standard suite of host cells used in our assays was giving us a representative picture of the maximum phage titre obtainable in a given sample, we used every cultured isolate of *Prochlorococcus* in our collection with a unique ITS rDNA sequence (23 isolates) to assay the cyanophage titre of a 50-m water sample from the Red Sea. The range of titres yielded was representative of the range we measured using our subset of *Prochlorococcus* isolates (see Supplementary Information).

Received 21 April; accepted 15 July 2003; doi:10.1038/nature01929.

- Liu, H., Nolla, H. A. & Campbell, L. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat. Microb. Ecol.* **12**, 39–47 (1997).
- Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
- Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: All the world’s a phage. *Proc. Natl Acad. Sci. USA* **96**, 2192–2197 (1999).
- Woese, C. R. Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396 (2000).
- Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
- Stent, G. S. *Molecular Biology of Bacterial Viruses* (W.H. Freeman, San Francisco, 1963).
- Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464–467 (1998).
- Rocap, G., Distel, D., Waterbury, J. B. & Chisholm, S. W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* **68**, 1180–1191 (2002).
- Urbach, E. & Chisholm, S. W. Genetic diversity in *Prochlorococcus* populations flow cytometrically sorted from the Sargasso Sea and Gulf Stream. *Limnol. Oceanogr.* **43**, 1615–1630 (1998).
- Waterbury, J. B. & Valois, F. W. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophage abundant in seawater. *Appl. Environ. Microbiol.* **59**, 3393–3399 (1993).
- Wichels, A. et al. Bacteriophage diversity in the North Sea. *Appl. Environ. Microbiol.* **64**, 4128–4133 (1998).
- Suttle, C. A. & Chan, A. M. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar. Ecol. Prog. Ser.* **92**, 99–109 (1993).
- Wilson, W. H., Joint, I. R., Carr, N. G. & Mann, N. H. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH 7803. *Appl. Environ. Microbiol.* **59**, 3736–3743 (1993).
- Lu, J., Chen, F. & Hodson, R. E. Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. *Appl. Environ. Microbiol.* **67**, 3285–3290 (2001).
- Murphy, F. A. et al. *Virus Taxonomy: Classification and Nomenclature of Viruses* (Springer, Vienna, Austria, 1995).
- Scanlan, D. J. & West, N. J. Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microb. Ecol.* **40**, 1–12 (2002).
- Suttle, C. A. & Chan, A. M. Dynamics and distribution of cyanophages and their effects on marine *Synechococcus* spp. *Appl. Environ. Microbiol.* **60**, 3167–3174 (1994).

18. Suttle, C. A. in *The Ecology of Cyanobacteria* (eds Whitton, B. A. & Potts, M.) 563–589 (Kluwer, Netherlands, 2000).
19. Fuller, N. J. *et al.* Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl. Environ. Microbiol.* **69**, 2430–2443 (2003).
20. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic ecosystems. *Limnol. Oceanogr.* **45**, 1320–1328 (2000).
21. Cavender-Bares, K. K., Karl, D. M. & Chisholm, S. W. Nutrient gradients in the western North Atlantic Ocean: Relationship to microbial community structure, and comparison to patterns in the Pacific Ocean. *Deep-Sea Res.* **48**, 2373–2395 (2001).
22. Mann, E. L., Ahlgren, N., Moffett, J. W. & Chisholm, S. W. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol. Oceanogr.* **47**, 976–988 (2002).
23. Waterbury, J. B., Watson, S. W., Valois, F. W. & Franks, D. G. Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Can. Bull. Fish. Aquat. Sci.* **214**, 71–120 (1986).
24. Bohannan, B. J. M. & Lenski, R. E. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol. Lett.* **3**, 362–377 (2000).
25. Steward, G., Smith, D. C. & Azam, F. Abundance and production of bacteria and viruses in the Bering and Chukchi Seas. *Mar. Ecol. Prog. Ser.* **131**, 287–300 (1996).
26. Rohwer, F. *et al.* The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol. Oceanogr.* **45**, 408–418 (2000).
27. Wilson, W. H., Carr, N. G. & Mann, N. H. The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH7803. *J. Phycol.* **32**, 506–516 (1996).
28. Rocop, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047 (2003).
29. Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424**, 1037–1042 (2003).
30. Moore, L. R. & Chisholm, S. W. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr.* **44**, 628–638 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the respective Captain and crew of the R/V *Oceanus* and the R/V *Endeavor* for their help on research cruises, which were made possible by an NSF grant to chief scientists J. Moffett and B. Binder, respectively. This work was funded in part by Grants from the Seaver Foundation, NSF and DOE to S.W.C., and an NIH Genome Training Grant, an MIT Research Fellowship and a WHOI Research Fellowship to M.B.S. We thank A. F. Post for Red Sea samples, G. Tan for technical assistance, J. A. King and P. Weigle for TEM suggestions, S. Casjens and F. Rohwer for help with prophage analysis, and W. H. Wilson for providing cyanophage isolates S-PM2 and S-WHM1.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.W.C. (chisholm@mit.edu).

Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem

T. S. Bibby¹, I. Mary², J. Nield¹, F. Partensky² & J. Barber¹

¹Wolfson Laboratories, Biochemistry Building, Department of Biological Sciences, South Kensington Campus, Imperial College London, London SW7 2AZ, UK

²Station Biologique, UMR 7127, Centre National de la Recherche Scientifique et Université Paris 6, BP74, F-29682 Roscoff Cedex, France

Prochlorococcus, the most abundant genus of photosynthetic organisms¹, owes its remarkably large depth distribution in the oceans to the occurrence of distinct genotypes adapted to either low- or high-light niches^{2,3}. The *pcb* genes, encoding the major chlorophyll-binding, light-harvesting antenna proteins in this genus⁴, are present in multiple copies in low-light strains but as a single copy in high-light strains⁵. The basis of this differentiation, however, has remained obscure. Here we show that the moderate low-light-adapted strain *Prochlorococcus* sp. MIT 9313 has one iron-stress-induced *pcb* gene encoding an antenna protein serving photosystem I (PSI)—comparable to *isiA* genes from cyanobacteria^{6,7}—and a constitutively expressed *pcb* gene encoding a photosystem II (PSII) antenna protein. By comparison, the very low-light-adapted strain SS120 has seven *pcb* genes

encoding constitutive PSI and PSII antennae, plus one PSI iron-regulated *pcb* gene, whereas the high-light-adapted strain MED4 has only a constitutive PSII antenna. Thus, it seems that the adaptation of *Prochlorococcus* to low light environments has triggered a multiplication and specialization of Pcb proteins comparable to that found for Cab proteins in plants and green algae⁸.

In order to gain a better understanding of the origin, function and localization of the divinyl-chlorophyll *a/b*-binding antenna complexes of *Prochlorococcus* species and how these properties relate to the light niche to which different strains are adapted, we have undertaken gene expression and structural studies on the moderate low-light-adapted strain MIT 9313 for which the full genome sequence is available (http://www.jgi.doe.gov/JGI_microbial/html/). This strain contains two *pcb* genes: *pcbA* (PMT 1046) and *pcbB* (PMT 0496). This contrasts with the very low-light-adapted strain SS120 that contains eight *pcb* genes (*pcbA* to *pcbH*)—analysis of its genome sequence (<http://www.sb-roscoff.fr/Phyto/ProSS120/>) revealed the presence of one more *pcb* gene (*pcbH*) than previously thought⁵—and the high-light-adapted strain MED4, which has only a single *pcb* gene (*pcbA*)^{4,5}.

Recently it was shown by electron microscopy and single-particle analyses that SS120 contains a giant supercomplex consisting of the PSII reaction centre trimer surrounded by a light-harvesting antenna ring composed of 18 Pcb subunits⁹, similar to the 18-mer IsiA–PSI supercomplex induced in cyanobacteria when deprived of iron^{6,7}. However studies on MIT 9313 grown under similar conditions to SS120 did not reveal the presence of an 18-mer Pcb–PSI supercomplex but only ‘naked’ trimeric PSI complexes, which matched with the cyanobacterial X-ray structure¹⁰ (Fig. 1a, b). Instead, electron microscopy (Figs 1c, d and 2a) indicated that Pcb proteins associate with the dimeric reaction centre complex of PSII to form a Pcb–PSII supercomplex having dimensions of approximately 210 × 290 Å. Our interpretation is that this Pcb–PSII supercomplex consists of eight Pcb subunits with four distributed on each side of the PSII dimer as shown in Fig. 1c. This is emphasized by overlaying onto the projection map the published X-ray-derived models of the PSII reaction centre dimer and CP43, a PSII antenna protein structurally similar to Pcb^{9,11} (Fig. 1d). In some cases, the four Pcb subunits on one side of the dimer were missing (Fig. 1e, f). The ‘naked’ PSI trimers and Pcb–PSII complexes shown in Fig. 1 were located in a chlorophyll-containing band (band 2 in Fig. 2b, insert +Fe) obtained by sucrose density centrifugation after solubilizing isolated thylakoid membranes with the detergent β-D-dodecyl maltoside. Also contained in this band were some PSII reaction centre dimers free of Pcb proteins (Fig. 1g, h). Analysis of all discernible particles, taken from band 2 sample micrographs, resulted in 1,192 particles assigned to PSI and PSII, and gave a PSI:PSII ratio of about 2. We assume this to be indicative of the ratio in the intact thylakoid membrane given that most PSI and PSII particles were in band 2. Amino-terminal sequencing of the Pcb protein in band 2 from +Fe conditions showed it to be the product of the *pcbA* gene only, a result which was also found for the free-Pcb proteins in band 1 and for Pcb protein in thylakoid membranes (Fig. 3).

The absence of the PcbB protein and the 18-mer Pcb–PSI supercomplex in iron-replete MIT 9313 cells spurred us to investigate the expression of *pcbA* and *pcbB* genes in this strain. When the cells were grown in medium supplemented with iron (+Fe), we found that only the *pcbA* gene was expressed (Table 1). However, when cells were transferred to culture medium without added iron (–Fe), expression of the *pcbB* gene was activated, a surprising result as *pcb* genes of *Prochlorococcus* were not known to be regulated by iron as is the iron-stress-induced *isiB* gene (Table 1). The latter gene encodes flavodoxin¹², which substitutes for ferredoxin as an electron acceptor to PSI, and its expression indicates that the cells had acclimatized to conditions of iron depletion. On the other hand, the expression of

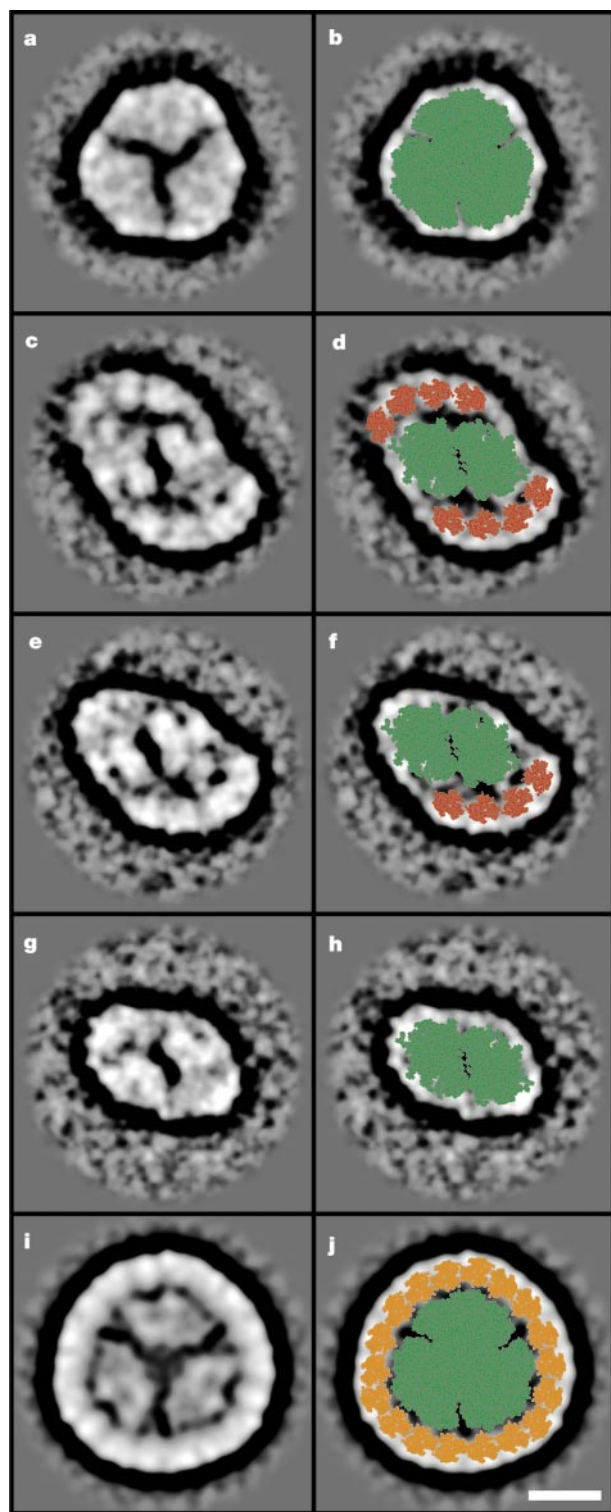


Figure 1 Characteristic top view of negatively stained particles isolated from *Prochlorococcus* MIT 9313. **a–j**, Particles isolated from cells grown in the presence (**a**, **c**, **e**, **g**) or absence (**i**) of iron viewed by electron microscopy and shown after overlaying the X-ray structures of the PSI trimer (green, **b**, **j**)¹¹, PSII core dimer (green, **d**, **f**, **h**) and CP43 (red/orange)¹². The Pcb proteins are coloured red or orange to identify them as products of the *pcbA* (PMT 1046) or *pcbB* (PMT 0496) genes respectively (see text). Scale bar, 100 Å.

the *pcbA* as well as the *psaA* and *psbA* genes, encoding PSI and PSII reaction centre proteins, respectively, were downregulated (Table 1).

The question arising from these expression studies is: where is the PcbB protein targeted in cells exposed to low iron levels? As before,

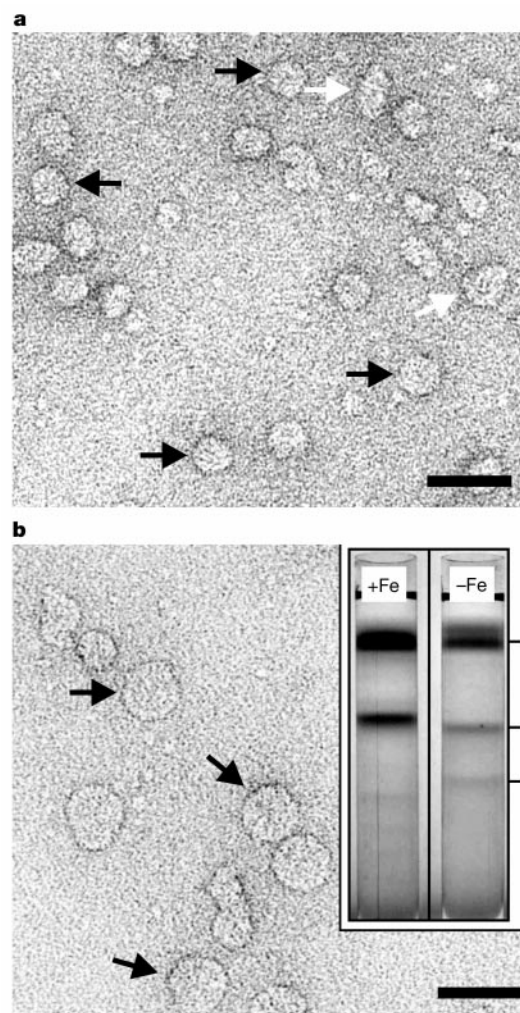


Figure 2 Electron micrographs of particles isolated from *Prochlorococcus* MIT 9313 separated on sucrose density gradients. **a**, **b**, Band 2 (**a**) of cells grown in the presence of iron (+Fe) containing both PSII particles (white arrow) and naked PSI trimers (black arrow), and band 3 (**b**) of the sucrose density gradient from cells grown in iron-deficient medium (–Fe) containing the 18-mer Pcb–PSI supercomplex (black arrow). The insert is the banding profile obtained by sucrose density gradient centrifugation of thylakoid membranes solubilized with 1% β -D-dodecyl maltoside isolated from *Prochlorococcus* MIT 9313. Band 1 consists of free Pcb proteins. Scale bar, 500 Å.

thylakoid membranes were isolated, solubilized and subjected to sucrose density gradient centrifugation. As shown in insert –Fe of Fig. 2b, an additional chlorophyll-containing band was observed compared with the iron-supplemented cells (band 3). Electron microscopy (Fig. 2b) revealed that this band contained the 18-mer Pcb–PSI supercomplex (Fig. 1i, j), similar to that observed in *Prochlorococcus* SS120 (ref. 9). SDS–polyacrylamide gel electrophoresis (PAGE) and N-terminal sequencing indicated that the Pcb protein of the PSI supercomplex is derived from the *pcbB* gene. This gene product was readily observed in SDS–PAGE profiles of thylakoid membranes isolated from cells grown under iron deficiency and shown in Fig. 3 (white asterisk). It ran at a slightly higher apparent molecular mass compared with the PcbA protein observed in both +Fe and –Fe cells (black asterisk), consistent with the difference in their predicted molecular masses of 38,511 Da (PcbA) and 40,737 Da (PcbB). Notably, despite the downregulation of the *pcbA* gene under iron-depleted conditions, a relatively high level of the PcbA protein was detected by SDS–PAGE (see Fig. 3) and Pcb–PSII structures similar to those present in band 2 of iron-

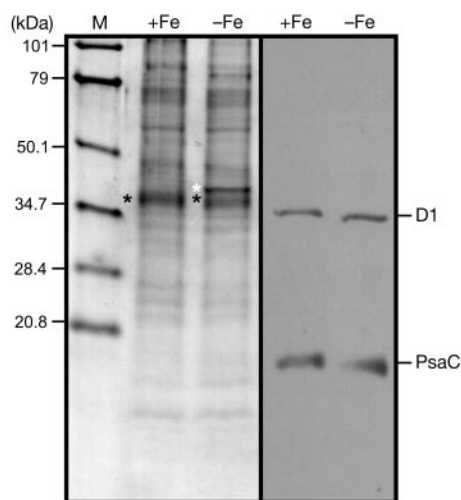


Figure 3 SDS-PAGE (left panel) and western-blot analyses (right panel) of thylakoid membranes isolated from *Prochlorococcus* MIT 9313 cells grown in the presence (+Fe) and absence (–Fe) of iron. The expression of the *pcbB* (PMT 0496) gene under –Fe conditions gives rise to an additional band in the SDS-PAGE gel, where PcbA and PcbB proteins are marked by black and white asterisks, respectively. The identity of these proteins was confirmed by N-terminal sequencing giving MQTYGNPN for PcbA and MQTYGKTD for PcbB. The antibodies for detecting the level of PSII were raised against PsbA (D1 protein) and for PSI raised against PsbA (F_A/F_B protein). The blotting indicates that the PSI:PSII ratio remains about the same for +Fe and –Fe cells.

supplemented (see Fig. 2, insert +Fe) cells were observed by electron microscopy and image analysis in iron-depleted cells. Intriguingly, although in cyanobacteria the induction of the 18-mer IsiA–PSI supercomplex under iron deficiency seems to compensate for the reduction in the level of PSI relative to PSII and for a decrease in the synthesis of phycobiliproteins, these reasons cannot apply to *Prochlorococcus* MIT 9313. Immunoblotting analyses (Fig. 3) showed that the PSI:PSII ratio, estimated to be about two, remained approximately the same in +Fe and –Fe conditions despite the lowering of the expression of *psaA* and *psbA* genes in iron-depleted cells (Table 1).

Phylogenetic analyses of the *pcb/isiA/psbB/psbC* gene superfamily^{5,13} indicate that the *pcbA* and *pcbB* genes of MIT 9313 partition into two different clusters and that the latter is very closely related to the *pcbG* gene of *Prochlorococcus* SS120, which is the gene providing the Pcb protein of the 18-mer Pcb–PSI supercomplex of this strain under iron-replete conditions, as indicated by SDS-PAGE and N-terminal sequencing (data not shown). In the case of MED4 we have been unable to detect an 18-mer *pcb*–PSI supercomplex either under iron-rich or iron-depleted conditions. Indeed in this strain the expression of the single *pcb* gene (*pcbA*) was not significantly effected by iron depletion (Table 1). In fact, using electron microscopy and associated image analyses of single particles we found that the Pcb proteins of MED4 associated with PSII in a similar fashion to those of MIT 9313.

Therefore we conclude that the PcbA protein of MIT 9313, similar to the PcbA protein of MED4, is targeted to PSII where they interact with the reaction centre dimer and increase the light-harvesting capacity of this photosystem. In contrast the PcbB protein of MIT 9313, similar to the PcbG protein of SS120, is targeted to PSI where it forms an 18-mer light-harvesting antenna ring around the PSI reaction centre trimer.

Under iron depletion conditions, expression of the *pcbC* gene of strain SS120 strongly increased while expression of *pcbG* was downregulated (Table 1). It is therefore possible that PcbC replaces PcbG in Pcb–PSI supercomplexes under these conditions.

The observations reported here give support to the hypothesis

Table 1 Quantitative RT-PCR analyses of the expression of the indicated genes under +Fe and –Fe conditions

<i>Prochlorococcus</i> strain	Gene	Induction ratio (±s.d.)
MIT 9313	<i>psaA</i>	0.22 ± 0.08
	<i>psbA</i>	0.20 ± 0.02
	<i>isiB</i>	14.72 ± 2.90
	<i>pcbA</i>	0.07 ± 0.01
	<i>pcbB</i>	7.36 ± 1.70
SS120	<i>psaA</i>	0.32 ± 0.14
	<i>psbA</i>	0.49 ± 0.08
	<i>isiB</i>	12.13 ± 3.66
	<i>pcbA</i>	0.51 ± 0.03
	<i>pcbB</i>	0.73 ± 0.10
	<i>pcbC</i>	23.92 ± 5.85
	<i>pcbD</i>	0.51 ± 0.18
	<i>pcbE</i>	0.37 ± 0.09
	<i>pcbF</i>	0.34 ± 0.04
	<i>pcbG</i>	0.28 ± 0.06
MED4	<i>pcbH</i>	0.86 ± 0.12
	<i>psaA</i>	0.89 ± 0.12
	<i>psbA</i>	0.67 ± 0.20
	<i>isiB</i>	7.94 ± 1.23
	<i>pcbA</i>	0.93 ± 0.09

These genes encode the PsaA subunit of PSI, the D1 subunit of PSII, flavodoxin and the chlorophyll *a/b*-binding light-harvesting complexes, respectively. Data were normalized to the expression of the housekeeping gene *mpb*, encoding RNase P. The ratio (the relative change in gene expression in –Fe compared with +Fe conditions) was calculated using the $\Delta\Delta C_T$ method as detailed by the manufacturer (http://dna-9.int-med.uiowa.edu/RealtimePCRdocs/Compar_Anal_Bulletin2.pdf) using the equation $(2)^{-\Delta\Delta C_T} = 2^{-(C_{T(\text{target gene} - \text{Fe})} - C_{T(\text{mpb} - \text{Fe})}) - (C_{T(\text{target gene} + \text{Fe})} - C_{T(\text{mpb} + \text{Fe})})}$ where C_T is the threshold cycle for amplification of target gene or *mpb* from –Fe or +Fe cultures.

that the common ancestor of *Prochlorococcus* and marine *Synechococcus* possessed an *isiA*-like gene, similar to that found in other cyanobacteria¹², even if, surprisingly, no *isiA* is present in the only currently available genome of a marine *Synechococcus* (strain WH8102; http://www.jgi.doe.gov/JGI_microbial/html/). The difference in Pcb structure between strains can be interpreted in terms of their respective light niche. For SS120, having both constitutive PSI and PSII antennae possibly confers to this strain an adaptive advantage to grow under the very low irradiances found at the bottom of the euphotic zone³. In the high-light-adapted strain MED4, only a PSII-related *pcb* gene is required because in the upper, well-illuminated layer of the ocean an additional antenna for PSI may not be needed. Indeed, according to the recent X-ray structure¹⁰, the ‘naked’ cyanobacterial PSI reaction centre already binds almost 100 light-harvesting chlorophyll molecules. The intermediate situation found in MIT 9313 is harder to interpret, as it appears to have an antenna ring around PSI only under iron depletion. Perhaps this suggests that there is a more complex inter-relationship between light intensity and availability of iron in the oceanic environment than hitherto considered.

The apparent requirement of at least some Pcb proteins to associate with PSII in *Prochlorococcus* is in line with the relatively low level of about 30 light-harvesting chlorophylls bound to PSII (ref. 12). Therefore in all types of photosynthetic organisms the light-harvesting capacity of PSII usually needs to be boosted by additional antenna systems: chlorophyll *a/b*-binding Cab proteins of plants and green algae⁸ and phycobiliproteins of cyanobacteria and red algae¹⁴. In the case of *Prochlorococcus*, each Pcb subunit adds an additional 13 chlorophylls assuming that they bind the same level of this pigment as CP43 (ref. 12). Consequently it takes the binding of just two Pcb subunits to approximately double the light-harvesting capacity of PSII. □

Methods

Isolation and biochemical characterization

Prochlorococcus sp. MIT 9313 (ref. 2) was grown in PCR-S11 medium¹. Iron-stressed cultures were obtained by two consecutive transfers of the cells into fresh medium containing only 1% of the standard FeCl₃ concentration of the PCR-S11 medium but having the normal Na₂-EDTA concentration. Cells in the exponential phase of growth were collected and thylakoid membranes isolated according to ref. 15. The thylakoids

(1 mg chlorophyll ml⁻¹) were then solubilized with 1% β -D-dodecyl maltoside at 4 °C for 10 min. The solubilized membranes were centrifuged at 100,000g using a Beckman Ti70 rotor and the solubilized fraction subjected to sucrose density centrifugation⁶. The resulting bands were independently removed for biochemical and structural characterization. SDS-PAGE and western blotting were performed according to ref. 16. Antibodies were a gift from P. Nixon (anti-D1) and J. Golbeck (anti-PsaC). N-terminal sequencing of proteins was performed by J. Keen.

RNA isolation and RT-PCR analysis

RNA was isolated from frozen cell pellets as previously described¹⁷. Quantitative RT-PCR reactions were performed on an ABI Prism 5700-sequence detection system (PE Applied Biosystems) as detailed elsewhere¹⁸ using specific primers of selected photosynthetic genes (see Table 1), defined according to their sequences in MIT 9313, MED4 and SS120 genome data banks. For each quantitative RT-PCR experiment, measurements were made in duplicate and experiments were repeated on two distinct cultures.

Electron microscopy

Isolated preparations were placed onto carbon-evaporated, glow-discharged, 300-mesh copper grids and negatively stained with 2% uranyl acetate and imaged using a Philips CM 100 electron microscope at 80 kV. The magnification was calibrated to be $\times 50,850$. Five electron micrographs were taken for each preparation where the first minima of the contrast transfer function was calculated as being in the range 21–22 Å. Micrographs were digitized using a Leafscan 45 densitometer at a step size of 10 μ m. All image processing was performed within the IMAGIC-5 software environment¹⁹ at a sampling frequency of 3.92 Å per pixel on the specimen scale. Single-particle data sets of approximately 1,200 (Pcb-PSI supercomplex; – Fe, band 3), 5,900 (+Fe, band 2) and 1,600 (– Fe, band 2) were obtained by selecting all possible single particles from the micrographs. Reference-free alignment and multi-variate statistical classification²⁰ were used to identify the initial class averages used for iterative refinement that resulted in the improved class averages shown.

Image processing

Coordinate data sets were obtained from the RCSB Data Bank (<http://www.rcsb.org>) under the entry codes for 1JB0 (PSI 2.5 Å structure¹⁰) and 1IZL (PSII 3.7 Å structure¹¹). These structural models were visualized using the program ViewerLite (Accelrys Inc.) and overlaid onto the calculated electron-microscopy-derived two-dimensional projection maps at the same scale. The coordinates of CP43 were extracted from the 1IZL.pdb file and modified by the removal of the large loop joining helices V and VI so as to better represent a typical Pcb protein. These coordinates were modelled into the centre of mass assigned to each Pcb subunit.

Received 3 April; accepted 15 July 2003; doi:10.1038/nature01933.

- Partensky, F., Hess, H. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
- Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464–467 (1998).
- Moore, L. R. & Chisholm, S. W. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr.* **44**, 628–638 (1999).
- La Roche, J. *et al.* Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light-harvesting proteins. *Proc. Natl Acad. Sci. USA* **93**, 15244–15248 (1996).
- Garczarek, L. *et al.* Expression and phylogeny of the multiple antenna genes of the low-light-adapted strain *Prochlorococcus marinus* SS120. *Plant Mol. Biol.* **46**, 683–693 (2001).
- Bibby, T. S., Nield, J. & Barber, J. Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* **412**, 743–745 (2001).
- Boekema, E. J. *et al.* A giant chlorophyll-protein complex induced by iron deficiency in cyanobacteria. *Nature* **412**, 745–748 (2001).
- Green, B. R. & Durnford, D. G. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 685–714 (1996).
- Bibby, T. S., Nield, J., Partensky, F. & Barber, J. Oxyphotobacteria: Antenna ring around PSI. *Nature* **413**, 590 (2001).
- Jordan, P. *et al.* Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).
- Kamiya, N. & Shen, J.-R. Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7 Å resolution. *Proc. Natl Acad. Sci. USA* **100**, 98–102 (2003).
- Strauss, N. A. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 731–750 (Kluwer, Dordrecht, The Netherlands, 1994).
- Partensky, F. & Garczarek, L. in *Photosynthesis in Algae. Advances in Photosynthesis Series* (ed. Larkum, A. W. D.) (Kluwer, Dordrecht, The Netherlands, in the press).
- Sidler, W. in *The Molecular Biology of Cyanobacteria* Vol. 14 (ed. Bryant, D. A.) 139–216 (Kluwer, Dordrecht, The Netherlands, 1994).
- Partensky, F. *et al.* The divinyl-chlorophyll *a/b*-protein complexes of two strains of the oxyphototrophic marine prokaryote *Prochlorococcus*—Characterization and response to changes in growth irradiance. *Photosynth. Res.* **51**, 209–222 (1997).
- Hankamer, B. *et al.* Isolation and biochemical characterization of monomeric and dimeric photosystem II complexes from spinach and their relevance to the organisation of photosystem II in vivo. *Eur. J. Biochem.* **243**, 422–429 (1997).
- Bhaya, D. *et al.* Isolation of regulated genes of the cyanobacterium *Synechocystis* sp. strain PCC 6803 by differential display. *J. Bacteriol.* **182**, 5692–5699 (2000).
- Mary, I. & Vaulot, D. Two-component systems in *Prochlorococcus marinus*: Genomic analysis and differential expression under stress. *FEMS Microbiol.* (in the press).
- van Heel, M., Harauz, G. & Orlova, E. V. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
- van Heel, M. & Frank, J. Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy* **6**, 187–194 (1981).

Acknowledgements J.B. thanks the Biotechnology and Biological Sciences Research Council for financial support. J.N. holds a University Research Fellowship of The Royal Society. F.P. and I.M. were supported by the EU FPS program Margenes, and I.M. was awarded a Ph.D. grant from Région Bretagne. We thank F. Le Gall for preparing large volumes of *Prochlorococcus* cultures.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.B. (j.barber@imperial.ac.uk).

High diversity of unknown picorna-like viruses in the sea

Alexander I. Culley¹, Andrew S. Lang^{2*} & Curtis A. Suttle^{1,2,3}

¹Department of Botany, ²Department of Earth and Ocean Sciences, ³Department of Microbiology and Immunology University of British Columbia, 1461-6270 University Boulevard, Vancouver, British Columbia, V6T 1Z4, Canada

* Present address: Institute of Marine Sciences, University of Alaska Fairbanks, Fairbanks, Alaska, 99775, USA

Picornalike viruses are a loosely defined group of positive-sense single-stranded RNA viruses that are major pathogens of animals, plants and insects. They include viruses that are of enormous economic and public-health concern and are responsible for animal diseases (such as poliomyelitis¹), plant diseases (such as sharka²) and insect diseases (such as sacbrood³). Viruses from the six divergent families (the Picornaviridae, Caliciviridae, Comoviridae, Sequiviridae, Dicistroviridae and Potyviridae) that comprise the picorna-like virus superfamily⁴ have the following features in common: a genome with a protein attached to the 5' end and no overlapping open reading frames, all the RNAs are translated into a polyprotein before processing, and a conserved RNA-dependent RNA polymerase (RdRp) protein. Analyses of RdRp sequences from these viruses produce phylogenies that are congruent with established picorna-like virus family assignments^{5–7}; hence, this gene is an excellent molecular marker for examining the diversity of picorna-like viruses in nature. Here we report, on the basis of analysis of RdRp sequences amplified from marine virus communities, that a diverse array of picorna-like viruses exists in the ocean. All of the sequences amplified were divergent from known picorna-like viruses, and fell within four monophyletic groups that probably belong to at least two new families. Moreover, we show that an isolate belonging to one of these groups is a lytic pathogen of *Heterosigma akashiwo*, a toxic-bloom-forming alga responsible for severe economic losses to the finfish aquaculture industry, suggesting that picorna-like viruses are important pathogens of marine phytoplankton.

Viruses are extremely abundant and geochemically significant agents of microbial mortality in the ocean^{8,9}. They comprise a morphologically and genetically diverse array of pathogens, some of which infect heterotrophic bacteria and cyanobacteria, as well as photosynthetic and non-photosynthetic protists^{10–12}. On the basis of a variety of evidence, marine viral communities have been assumed to consist almost entirely of double-stranded DNA viruses; consequently, little effort has been made to examine natural communities of RNA viruses. There are data however to indicate that marine RNA viruses are also important. For example, rhabdoviruses and paramyxoviruses are negative-sense, ssRNA viruses that are major pathogens of fish¹³ and marine mammals¹⁴, respectively. In addition, picorna-like viruses have been isolated that are pathogens of penaeid shrimp⁷, seals¹⁵ and whales¹⁵.

We used available sequences in GenBank to design degenerate primers that target the highly conserved RdRp sequence in picorna-like viruses. These primers in conjunction with RT-PCR were used to assay for the presence of picorna-like viruses in the coastal waters of British Columbia, Canada. With the exception of retroviruses, all RNA viruses encode an RdRp, which is essential for replication. Within the RdRp protein sequence, several motifs have been identified which are homologous among diverse species of RNA viruses^{5,16}. Groups based on alignments of these conserved regions are congruent with presently defined RNA virus families^{6,17}. Families of picorna-like viruses classified on the basis of RdRp sequence data are congruent with families of picorna-like viruses classified according to virus structure, host and epidemiology. Consequently, sequence analysis can be used to infer the relationship of sequences from natural viral communities to known families of picorna-like viruses.

Viruses were concentrated from sea water using ultrafiltration¹⁸ in the spring and summer of 1996, 1997 and 2000. Viral RdRp sequences were amplified from 13 of the 21 viral communities examined. Amplification occurred in samples collected from oceanographically diverse environments, including anthropogenically influenced sites, estuarine environments, stations with a well-mixed water column and pristine, highly stratified fjords. Preliminary analysis of the environmental sequences by BLAST¹⁹ searches of the GenBank database gave high similarities to several picorna-like virus family RdRp sequences as well as an unidentified viral sequence from a Chinese clam homogenate²⁰. These results suggested that picorna-like viruses are prevalent in a variety of marine waters.

To confirm the amplified products originated from picorna-like viruses, a selection of the amplicons were sequenced, and along with representative sequences from known picorna-like virus families,

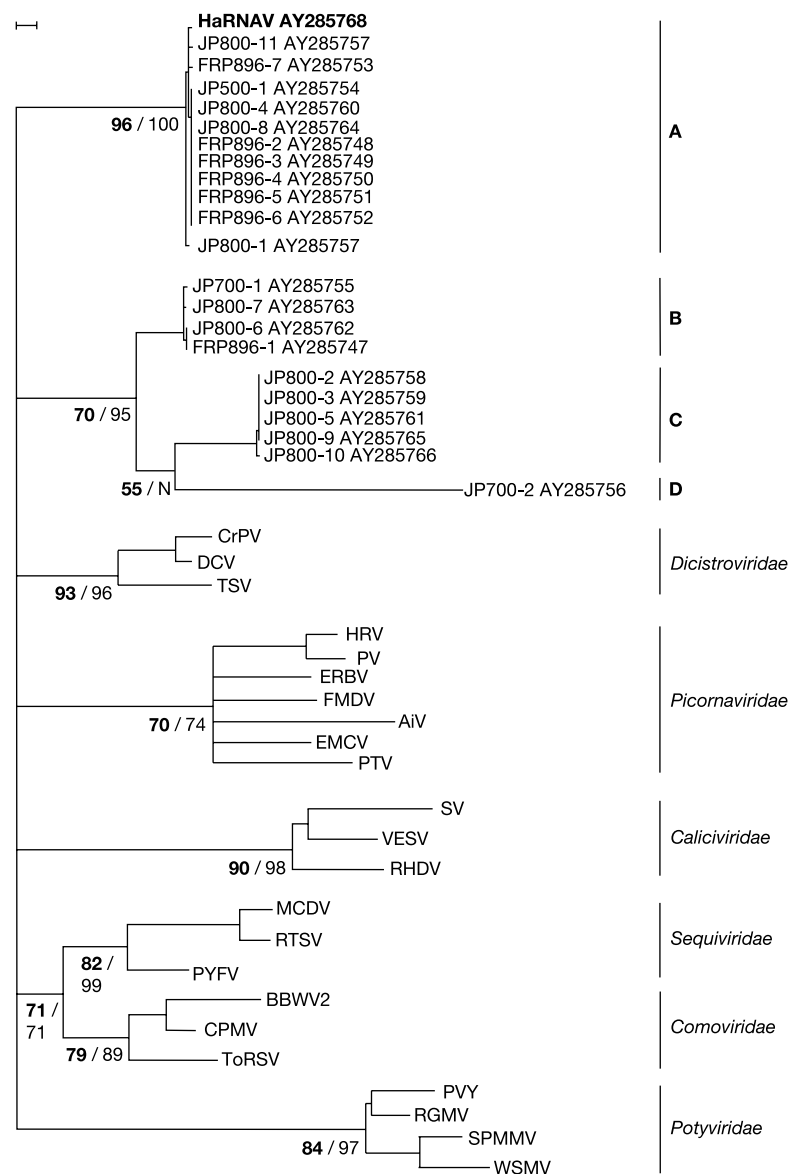


Figure 1 Maximum-likelihood tree of RdRp sequences from environmental amplicons and representative viruses from picorna-like virus families. (See Methods for complete virus names). Viruses from the Potyviridae, which contain RdRp sequences from a different lineage⁵, were used as an outgroup. Family names and group letters are shown. Environmental amplicons from coastal British Columbia are labelled by a two- or three-letter station designation, month, year, group sequence number and GenBank database

accession number (SSSMYY-AA, see also Supplementary Information). TREE-PUZZLE support values are shown for relevant nodes in boldface followed by bootstrap values based on neighbour-joining analysis. N indicates there was no corresponding node in the neighbour-joining tree. The maximum-likelihood distance scale bar indicates a distance of 0.1.

used to construct phylogenetic trees. All sequenced environmental PCR products translated into continuous amino acid sequences that contained the signature positive-stranded ssRNA virus RdRp motif GDD²¹. Phylogenetic analysis of the RdRp fragment amplified in this study resolved all established picorna-like families (Fig. 1). Strikingly, none of the environmental sequences fell within established families of picorna-like viruses, but rather into four previously unknown and distantly related groups that we refer to as A, B, C and D. Our results suggest that at least two and possibly all four of these groups represent new families of RNA viruses. Phylogenetic trees constructed with both maximum-likelihood and neighbour-joining methods group the environmental sequences outside established picorna-like virus families (Fig. 1). However, low bootstrap support prevents us from drawing any further conclusions regarding the evolutionary relationship between groups B, C and D. Interestingly, a single sample (JP800) collected from English Bay, which is adjacent to the Strait of Georgia and the city of Vancouver, contained representatives from three of the four novel clades, indicating that a high diversity of picorna-like viruses exists even within a single water sample.

Sequence identity among the clades of environmental picorna-like virus sequences was low, ranging from 38.9% to 54.6% nucleotide identity and 21.8% to 52.9% identity when translated to amino acids. In contrast, sequences within each clade were highly conserved and ranged from 97.7% to 100.0% and 95.9% to 100.0% identical on nucleotide and amino-acid levels, respectively. This is more variability than would be expected from potential errors in RT-PCR and sequencing; therefore, we are confident that despite the high identity within these groups, the sequences represent a number of different viruses. Moreover, within group A, although no sequences were identical on a nucleotide level, six were identical when translated into amino acids; this suggests that the observed nucleotide differences are real and not due to methodological error.

Although four groups of picorna-like viruses were discovered in this study, one of these groups (group A) contains a virus (HaRNAV²²) that causes the lysis of *Heterosigma akashiwo*, a toxic-bloom-forming alga that is responsible for major fish deaths in temperate waters²³. Nucleotide sequences from HaRNAV were 98.8% to 99.7% identical to sequences from group A, implying that there are a number of viruses closely related to HaRNAV that belong within this group. Interestingly, we were able to amplify sequences belonging to group A from samples collected in different locations, seasons and years, suggesting that HaRNAV-like viruses are reoccurring and widely distributed in the Strait of Georgia.

Our results indicate that there is a diverse but previously unknown community of picorna-like viruses that are persistently occurring and widespread in the ocean. The fact that sequences from four stations resulted in at least two novel families of picorna-like viruses suggests that the diversity of these viruses in the ocean is high. Branch lengths between groups B, C and D are similar to those among families and genera of known picorna-like viruses (Fig. 1), suggesting that these groups are representative of at least three novel genera within a new family, or, in fact, may represent three previously unknown families.

Viruses are obligate pathogens that generally remain infectious for a relatively short time in natural marine waters²⁴. The repeated amplification of picorna-like virus sequences from the same geographic location in water samples collected over a four-year period implies persistent viral production and therefore infection. The significant, high degree of identity between a sequence amplified from clams harvested from Asian waters and marine group B sequences from this study suggests that picorna-like viruses are likely to be present in a wide range of marine environments. Furthermore, the few isolates of marine picorna-like viruses infect ecologically and economically important organisms. These include HaRNAV²², which infects the red-tide-forming, fish-killing raphidophyte *Heterosigma akashiwo*, Taura syndrome virus (TSV), which

infects penaeid shrimp⁷, an intensely farmed, important member of the marine food web, and San Miguel sea lion viruses (SMSVs), which infect pinnipeds such as the California sea lion (*Zalophus californianus*)²⁵. Ultimately these newly identified viruses should be isolated, sequenced in full and their hosts identified. In the terrestrial environment, picorna-like viruses infect a wide variety of organisms and are responsible for several important diseases; it seems likely that they will prove to be important players in marine ecosystems as well. □

Methods

Sample collection and preparation

Viruses from 40 to 200 litres of sea water were concentrated from stations in and adjacent to the Strait of Georgia, British Columbia, between May and August during 1996, 1997 and 2000 aboard the CCGS *Vector* as described¹⁸. Two millilitres from each concentrated viral community was pelleted by ultracentrifugation and RNA extracted from resuspended pellets using Trizol-LS (Invitrogen) as per the manufacturer's protocol.

RT-PCR

The degenerate primers RdRp 1 (negative-strand, 5'-GGA/GGAC/TTACAG/CCIA/GA/TTTGTAT-3') and RdRp 2 (5'-A/CACCAACG/TA/CCG/TCTTG/CAA/GA/GAA-3') were designed from an alignment of the putative RdRp sequences from several picorna-like viruses in the GenBank database. To confirm the identity of environmental PCR products, the 454-base pair (bp) target fragment includes a highly conserved amino acid motif (GDD) characteristic of the RdRp of positive-strand RNA viruses²¹.

Complementary DNA was synthesized with SuperScript II RNaseH⁻ Reverse Transcriptase (Invitrogen) with the reagents provided using 5 µl of extracted RNA and the positive-strand primer RdRp 2. Subsequently, PCR was performed with RdRp 1 and RdRp 2 primers. PCR products were separated on a 1.5% agarose gel and bands of the appropriate size (approximately 500 bp) were excised. Washed agarose plugs were used as the template in a second round of PCR with the RdRp primer set. Positive and negative controls were done in parallel for the entire procedure. These RdRp fragments were either directly cloned or separated using denaturing gradient gel electrophoresis (DGGE).

DGGE was conducted using 25% to 40% linear denaturant gradient, 7% to 8% linear polyacrylamide gradient gels, as described²⁶. DGGE bands were excised, re-suspended and amplified in a third round of PCR with RdRp primers.

Cloning and sequencing

Second-round PCR fragments or re-amplified excised DGGE bands were cloned in the pGEM-T (Promega) vector using the manufacturer's protocol. Recombinants containing the cloned insert were identified using PCR with universal -21M13 (5'-GTTTCCCAGT CACGACGTTGTA-3') and M13R (5'-CAGGAACAGCTATGACC-3') primers. Cloned, second-round PCR products were screened by restriction endonuclease digestion. PCR products from plasmids with DGGE bands and second-round PCR inserts with unique digestion patterns were sequenced. PCR fragments were sequenced at the University of British Columbia Nucleic Acid and Protein Service Facility (Vancouver, Canada). Conserved regions of translated sequences were aligned with CLUSTAL X v1.81 (ref. 27) and then transformed into maximum-likelihood distances using the WAG matrix²⁸ in TREE-PUZZLE v5.0 (ref. 29) and 25,000 puzzling steps. Neighbour-joining bootstrap values were calculated based on 1,000 replicates using FITCH v3.6 (ref. 30). The name, acronym and accession number of viruses used in Fig. 1 are: Aichi virus (AIV), NC_001918; broad bean wilt virus 2 (BBWV2), AB013615; cowpea mosaic virus (CPMV), NC_003549; cricket paralysis virus (CrPV), NC_003924; *Drosophila* C virus (DCV), NC_001834; encephalomyocarditis virus (EMCV), NC_001479; equine rhinitis B virus (ERBV), NC_003983; foot-and-mouth disease virus (FMDV), NC_004004; human rhinovirus A (HRV), NC_001490; maize chlorotic dwarf virus (MCDV), NC_003626; parsnip yellow fleck virus (PYFV), NC_003628; poliovirus (PV), NC_002058; porcine teschovirus (PTV), NC_003985; potato virus Y (PVY), NC_001616; rabbit haemorrhagic disease virus (RHDV), NC_001543; rice tungro spherical virus (RTSV), NC_001632; ryegrass mosaic virus (RGMV), NC_001814; Sapporo virus (SV), U65427; sweet potato mild mottle virus (SPMMV), NC_003797; swine vesicular exanthema virus (VESV), NC_002551; Taura syndrome virus (TSV), NC_003005; tomato ringspot virus (ToRSV), NC_003840; wheat streak mosaic virus (WSMV), NC_001886.

We determined an error rate of 0.3% to 0.9% based on five independent amplifications using RT-PCR of a picorna-like virus isolate, implying that sequences less than 99.0% identical are probably different.

Received 7 May; accepted 7 July 2003; doi:10.1038/nature01886.

1. Pallansch, M. A. & Roos, R. P. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 723–775 (Lippincott, Williams & Wilkins, Philadelphia, 2001).
2. Lazarowitz, S. G. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 533–598 (Lippincott, Williams & Wilkins, Philadelphia, PA, 2001).
3. Bailey, L., Gibbs, A. J. & Woods, R. D. Sacbrood virus of the larval honey bee (*Apis mellifera* Linnaeus). *Virology* **23**, 425–429 (1964).
4. Liljas, L. et al. Evolutionary and taxonomic implications of conserved structural motifs between picornaviruses and insect picorna-like viruses. *Arch. Virol.* **147**, 59–84 (2002).
5. Koonin, E. V. & Dolja, V. V. Evolution and taxonomy of positive-strand RNA viruses - implications of comparative-analysis of amino-acid-sequences. *Crit. Rev. Biochem. Mol. Biol.* **28**, 375–430 (1993).
6. Zanotto, P. M. D., Gibbs, M. J., Gould, E. A. & Holmes, E. C. A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *J. Virol.* **70**, 6083–6096 (1996).

7. Mari, J., Poulos, B. T., Lightner, D. V. & Bonami, J. R. Shrimp Taura syndrome virus: genomic characterization and similarity with members of the genus *Cricket paralysis-like viruses*. *J. Gen. Virol.* **83**, 915–926 (2002).
8. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
9. Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea. *Bioscience* **49**, 781–788 (1999).
10. Suttle, C. A. in *Viral Ecology* (ed. Hurst, C. J.) 248–286 (Academic, San Diego, CA, 2000).
11. Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
12. Mann, N. H. Phages of the marine cyanobacterial picophytoplankton. *FEMS Microbiol. Rev.* **27**, 17–34 (2003).
13. Bernard, J. & Bremont, M. Molecular biology of fish viruses — a review. *Vet. Res.* **26**, 341–351 (1995).
14. Van Bresse, M. F., Van Waerebeek, K. & Raga, J. A. A review of virus infections of cetaceans and the potential impact of morbilliviruses, poxviruses and papillomaviruses on host population dynamics. *Dis. Aquat. Organ.* **38**, 53–65 (1999).
15. Smith, A. in *Viral Ecology* (ed. Hurst, C.) 447–491 (Academic, San Diego, CA, 2000).
16. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8**, 3867–3874 (1989).
17. Koonin, E. V. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* **72**, 2197–2206 (1991).
18. Suttle, C. A., Chan, A. M. & Cottrell, M. T. Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Appl. Environ. Microbiol.* **57**, 721–726 (1991).
19. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
20. Kingsley, D. H., Meade, G. K. & Richards, G. P. Detection of both hepatitis A virus and Norwalk-like virus in imported clams associated with food-borne illness. *Appl. Environ. Microbiol.* **68**, 3914–3918 (2002).
21. Kamer, G. & Argos, P. Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12**, 7269–7282 (1984).
22. Tai, V. *et al.* Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Raphidophyceae). *J. Phycol.* **39**, 343–352 (2003).
23. Smayda, T. J. in *Physiological Ecology of Harmful Algal Blooms* (eds Anderson, D. M., Cembella, A. D. & Hallegraeff, G. M.) 113–131 (Springer, Berlin, GER, 1998).
24. Suttle, C. A. & Chen, F. Mechanisms and rates of decay of marine viruses in seawater. *Appl. Environ. Microbiol.* **58**, 3721–3729 (1992).
25. Smith, A. W., Akers, T. G., Madiu, S. H. & Vedros, N. A. San Miguel sea lion virus isolation, preliminary characterization and relationship to vesicular exanthema of swine virus. *Nature* **244**, 108–110 (1973).
26. Short, S. M. & Suttle, C. A. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl. Environ. Microbiol.* **68**, 1290–1296 (2002).
27. Thompson, J. D. *et al.* The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876–4882 (1997).
28. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
29. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
30. Fitch, W. M. & Margoliash, E. Construction of phylogenetic trees. *Science* **155**, 279–284 (1967).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the crew and participants of the SOG cruises from 1996 to 1998. We also thank C. Frederickson for assistance with DGGE and M. Berbee and P. Keeling for advice on phylogenetic analyses. This work was supported by a National Science and Engineering Research Council of Canada grant to C.A.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.A.S. (csuttle@eos.ubc.ca).

Mechanism of silk processing in insects and spiders

Hyoung-Joon Jin & David L. Kaplan

Departments of Chemical & Biological Engineering & Biomedical Engineering, Tufts University, Medford, Massachusetts 02155, USA

Silk spinning by insects and spiders leads to the formation of fibres that exhibit high strength and toughness¹. The lack of understanding of the protein processing in silk glands has prevented the recapitulation of these properties *in vitro* from reconstituted or genetically engineered silks. Here we report the identification of emulsion formation and micellar structures from aqueous solutions of reconstituted silkworm silk fibroin

as a first step in the process to control water and protein–protein interactions. The sizes (100–200 nm diameter) of these structures could be predicted from hydrophobicity plots of silk protein primary sequence². These micelles subsequently aggregated into larger ‘globules’ and gel-like states as the concentration of silk fibroin increased, while maintaining solubility owing to the hydrophilic regions of the protein interspersed among the larger hydrophobic regions. Upon physical shearing or stretching structural transitions, increased birefringence and morphological alignment were demonstrated, indicating that this process mimics the behaviour of similar native silk proteins *in vivo*. Final morphological features of these silk materials are similar to those observed in native silkworm fibres.

Many types of silk have been characterized, including cocoon silk fibroin from the silkworm, *Bombyx mori*, and dragline silk from the spider *Nephila clavipes*. Silkworm cocoon silk contains two structural proteins, the fibroin heavy (~350 kDa) and light (~25 kDa) chains, together with a family of sericin proteins to glue the fibroin brins together in forming the cocoon¹. The sericins are hydrophilic, and easily removed from cocoons by boiling in water³. The amino-acid sequence of the β -sheet forming crystalline region of fibroin is dominated by the hydrophobic sequence GAGAGSGAAG[SG(AG)₂]₈Y (refs 4, 5). Silk fibre formation involves shear and elongational stress acting on the fibroin solution (up to 30% wt/vol.) in the gland, causing fibroin in solution to crystallize¹. The process involves a lyotropic liquid crystal phase,

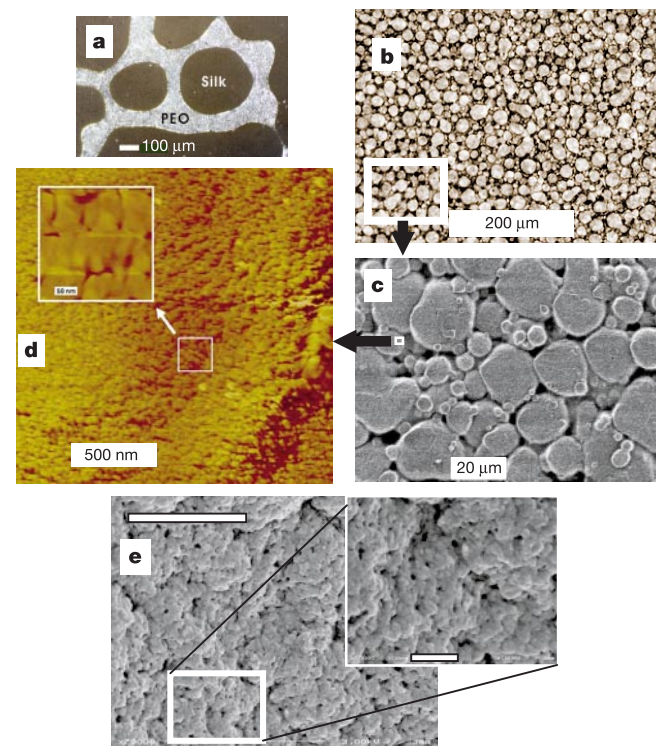


Figure 1 Optical polarizing microscope image of silk/PEO blend film cast from aqueous solution at ambient conditions. **a**, Silk/PEO (95/5 wt%) blend showing large ‘globules’. Dark regions belong to the silk fibroin phase, and bright regions to the continuous PEO phase. **b**, Silk/PEO (80/20 wt%) blend showing smaller ‘globules’. **c**, Examination of the indicated part of **b** by SEM, illustrating globule substructures (micelles) after the PEO was extracted with water. **d**, AFM of individual micelles in the 80/20 blend (small window scale bar, 50 nm). **e**, SEM image of silk solution collected from the posterior region of the *B. mori* gland, illustrating micellar-like morphology (gland contents were separated in water and treated in methanol; scale bars are 2 μ m on left, and 500 nm on right).

which is transformed from a gel to a sol state during spinning—that is, a liquid crystal spinning process¹. Elongational flow orients the fibroin chains, and the liquid is converted into filaments.

Degumming (deserincization) of cocoons generates sericin-free fibroin fibres that can be subsequently solubilized in concentrated aqueous salts such as LiBr (refs 5, 6) and formed into aqueous salt-free solutions by dialysis. In our recent studies with these aqueous fibroin solutions aimed at mimicking silk processing, we substituted polyethylene oxide (PEO) for the sericin⁷. We examined the solution behaviour of these aqueous fibroin/PEO blends (Fig. 1), and observed 'globules' with sizes that varied depending on the content of PEO (100–200 μm diameter at 5% PEO, <1–15 μm at 20% PEO). Closer examination of these globules by scanning electron microscopy (SEM) and atomic force microscopy (AFM) indicated a substructure morphology suggestive of micelles that were ~100–200 nm in diameter (Fig. 1).

On examination of the primary sequence for the fibroin heavy chain, a pattern of hydrophobic and hydrophilic blocks was identified that includes larger hydrophilic blocks at the chain ends, six smaller hydrophilic internal blocks and seven internal hydrophobic blocks (Fig. 2) (refs 2, 4–8). This pattern suggested the possibility of formation of micellar structures in water (Fig. 2). The smaller hydrophilic blocks may remain hydrated initially during fibroin self-assembly, and the larger terminal hydrophilic blocks at both the amino and carboxy termini define the outer edges of the micelles. The interpretation of this behaviour is supported by data on regenerated silk that is water-soluble just after dialysis from lithium bromide (ref. 9); the protein initially has a more extended chain structure until the hydrophobic regions begin to assemble and organize into micelles, and eventually gel. This process slows at

colder temperature, where complete hydrogelation of an 8% silk fibroin solution requires 5 ± 1 days at room temperature compared with 40 ± 5 days at 7 °C.

This model of silk assembly suggests that fibroin molecules act as hydrophilic–hydrophobic–hydrophilic polymers, with the formation of irregular sized micelles depending on chain folding and hydrophobic interactions (Fig. 2). There are, however, some important distinctions between these protein polymers and traditional synthetic triblocks^{10–12}. The subtle intervening hydrophilic blocks internal among the hydrophobic blocks in fibroin promote solubility in water for these micelles, and, more importantly, prevent premature β -sheet formation. In addition, there are extensive interactions among the blocks (hydrogen bonding, hydrophobic interactions) and interactions between the chain-end larger hydrophilic blocks. Each of these sets of interactions plays a critical role in the solubility of the fibroin solution, and avoidance of crystallization at this stage in the process. As the concentration of fibroin increases, inter-micelle interactions increase, leading to coalescence and the formation of globules and then gel states (as is seen in the glands upon dissection). The formation of nematic phases is a first step in the organization of the protein, and may occur prior to or concurrent with micelle formation¹. The fibroin solutions used in the present study are initially significantly more dilute than in the middle region of the gland during silk processing in nature; nonetheless, micellar morphology can be seen even in silkworm glands after dissection (Fig. 1).

To further extend this process, we determined structural features of fibroin films cast from these fibroin aqueous solutions, both with and without PEO. The films were subsequently treated with methanol or physical shear or stretching (1.5 times in length) (Fig. 3). The

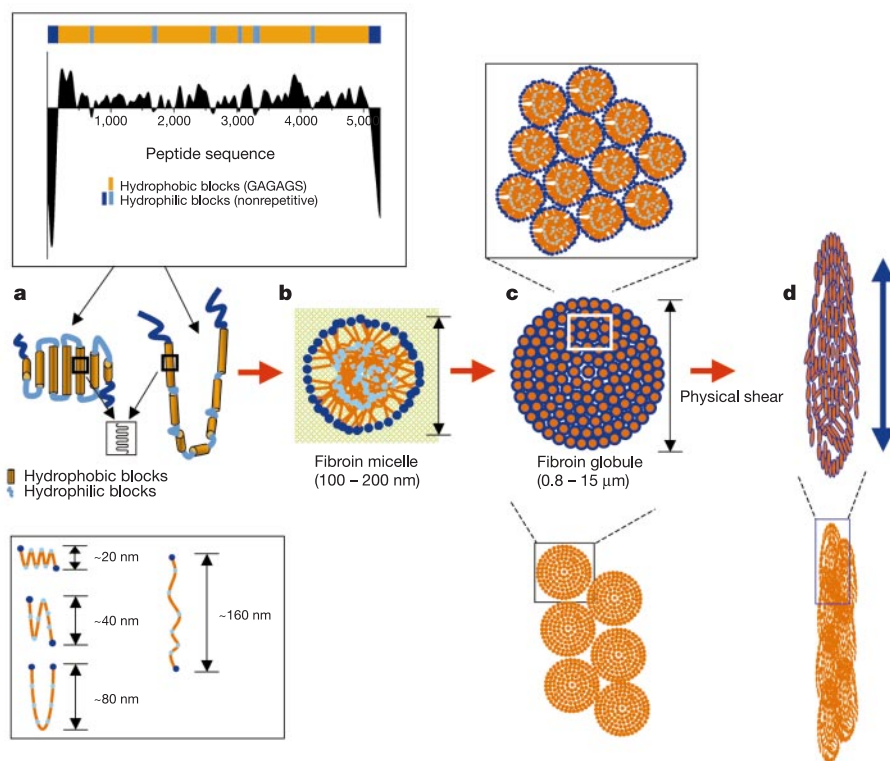


Figure 2 Model of chain folding, micelle formation, globule formation and curing, and shear processing of silk proteins. **a**, Hydrophobicity pattern in *B. mori* silk fibroin heavy chain primary sequence with possible chain folding intra- and inter-molecular schemes. **b**, Micelle assembly of silk fibroin in water (8%), based on hydrophilic–hydrophobic multiblock co-polymer structure leaving internal smaller hydrophilic domains to promote

solubility in water, with larger-chain terminal hydrophilic blocks in contact with the surrounding aqueous solution. **c**, 'Globule' formation driven by increased fibroin concentration and lower water content, further hydrophobic interactions, and at the final stages by the presence of sericin or PEO. **d**, Elongation and alignment of globules and interactions among globules promoted by physical shear, leading to fibrillar structure.

cast films (pure fibroin or blends with PEO) were stable in water and swelled, forming hydrogels. The silk fibroin films were characterized by wide-angle X-ray diffraction; the results indicated that a silk I structure predominated in the cast films before methanol treatment, while mixed structures (silk I, silk II, and undefined) were obtained by exposure to methanol or by stretching. Three silk fibroin conformations have been previously identified by X-ray diffraction and infrared spectroscopy; random coil (low concentrations of fibroin), α -form (silk I, type II β -turn, high concentrations formed in the absence of physical shear), and β -form (silk II, antiparallel β -pleated sheet, formed with physical shear or exposure to solvents such as methanol) (refs 13–19). The corresponding d -spacings for silk I and II are^{13–19} (in angstroms): 9.8 (II), 7.4 (I), 5.6 (I), 4.8 (II), 4.4 (I), 4.3 (II), 4.1 (I), 3.6 (I), 3.2 (I), 2.8 (I).

Upon stretching, the hydrogel film formed a dried solid film by expelling water or PEO that visually accumulated on the surface. The stretched films showed enhanced birefringence, as well as preliminary alignment of the globules along the line of applied shear (Fig. 3). Evidence for fibrillar substructures in native silk fibres has been reported^{20,21}, and in our recent studies on controlled

extraction of these fibres we could easily visualize the separation of these 8–10- μ m-diameter individual fibrils³. Furthermore, fracture surfaces of native silk fibres often display internal or core globular structures and a more filament-like coat^{22,23} (Fig. 4), consistent with our observations of this morphology in the films and with fibres spun from the pure fibroin gels, even in the absence of any direct mixing with PEO or sericin (Fig. 4). Studies on liposome formation with hydrophilic and hydrophobic polymer mixtures to induce micelle formation can explain some of the larger-scale phase separation phenomena observed^{24–26}; but because the same morphology appears even in the absence of direct additions of the PEO to the fibroin, the fundamental phenomena appear to be directed by the fibroin itself, as outlined above.

An explanation for this phase behaviour related to silk processing may rest first with hydrophobic/hydrophilic partitioning and chain folding. Subsequent transitions are driven by increases in fibroin concentration (due to biosynthesis as well as active transport of water from the gland), leading to well-described liquid crystalline states^{1,27,28}. In the final stages of spinning, absorption of water from the fibroin by PEO (or in the case of native silk processing, upon

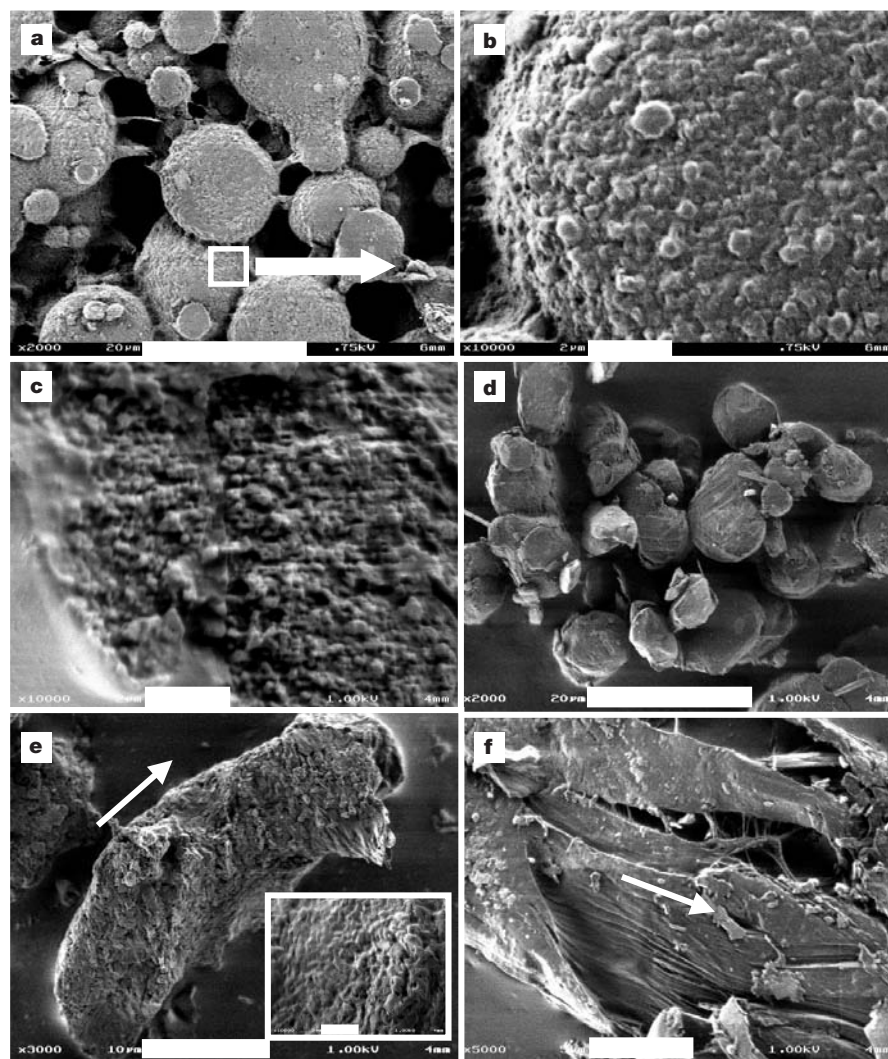


Figure 3 Scanning electron micrographs of silk/PEO (80/20 wt%) blend films after PEO extraction in water at room temperature for 24 h. **a**, Globule structures at low magnification (scale bar, 20 μ m). **b**, Globule structures at higher magnification (scale bar, 2 μ m) illustrating the subglobule morphology with $\sim$ 100-nm micellar morphology. **c**, Fracture surface of native silk fibre (scale bar, 2 μ m) after tensile failure at a rate of

20 mm min⁻¹, illustrating micelle substructure. **d**, Non-stretched globules (scale bar, 20 μ m). **e**, Intermediate coagulation/fusion of globules by shearing (scale bar, 10 μ m) along the direction of applied shear (arrow). Inset, 2 μ m. **f**, Aligned globules (scale bar, 5 μ m) along the direction of applied shear (arrow), illustrating beginning fibrillar-like morphology.

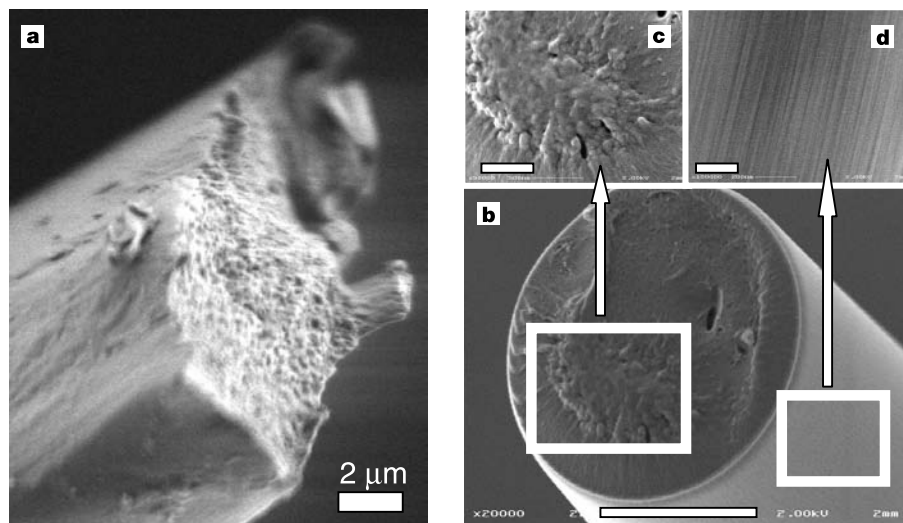


Figure 4 Fractured fibres. **a**, Scanning electron micrograph of fractured native silk fibroin after sericin extraction and tensile failure at a rate of 20 mm min^{-1} , illustrating internal micelle substructure and skin coat of aligned fibrillar structure. **b**, SEM of fibre

spun from pure fibroin solution ($\sim 16\%$ by weight fibroin) in the absence of PEO or sericin, illustrating core micellar/globular morphology (**c**) as well as skin coat (**d**) of aligned protein. Scale bars: **b**, $2 \mu\text{m}$; **c**, 500 nm ; **d**, 200 nm .

introduction of sericin) would result in an increase in local fibroin concentration due to the loss of water from the micelles to the hydrophilic phase, and induce the formation of the silk I structure. The globules now would no longer be able to redissolve in water even in the absence of exposure to alcohol, supporting a well-organized yet still soluble structure due to the internal hydrophilic blocks with entrained water.

Hydrophobic patterning (blocks) directs the folding and formation of micelles, and determines the phase behaviour of the structural proteins in silks. Subsequent coalescence of the micelles occurs owing to the continued increase in protein concentration. Final transitions are initiated by the addition of a second polymer (sericin) in the case of the silkworm, PEO in the case of mimicking the process as reported here, and subsequently by physical shear in the spinning step. In the case of sericin and PEO, these hydrophilic polymeric additions drive changes in globule size at the last stages in the process, and may accelerate the further assembly of fibroin into a silk I structure as well as define the final morphology of the fibres (sericin remains as a separate phase surrounding the fibroin core fibres in spun silkworm silk fibres). As indicated, these additions are not however required, as the process can be recapitulated in the absence of added PEO (Fig. 4 right). In the case of the major protein from the dragline (major ampullate gland silk) of *N. clavipes*, the hydrophilic/hydrophobic regions of the sequence (only a partial C-terminal sequence is available) are much smaller than in the silkworm fibroin. Further assembly into the silk I structure may be driven by the transport of water from the gland. The process described can also account for the formation of subfibrillar morphology reported for spider dragline silk since the globules can align due to the physical shear^{13,20,29} (Fig. 2).

In nature, water is the plasticizer and lubricant, and its entrainment in the micelles is critical to avoid premature crystallization in the initial and intermediate stages of silk processing. The process of avoiding crystallization until silk spinning may be further moderated by the presence of additional proteins, or perhaps biochemical changes in the gland. Final steps in the process (including water loss from the micelles) can be accounted for by the addition of sericin, by adding PEO in the case of the present study, or by further hydrophobic interactions or assembly of the protein in the case of the dragline silk of the spider. This intermediate lyotropic stage in silk processing may reflect the variation in silk I-like structures

reported in the literature. In the final stage of the process, physical shear elongates and aligns the globule-containing micelles and initiates the final structural transition to the insoluble crystalline β -sheet (silk II) to establish the basis for the high-strength and tough silk fibres. The shear forces induce interactions among the elongated globules, probably through limited hydrophobic interactions, generating more robust fibres. The ability to recapitulate silk processing *in vitro* suggests new options in aqueous processing of these types of protein, as well as polymers in general. However, we note that the subtle interactions among protein blocks, as well as in the sizes and spacing of these blocks, play a critical role in the process, and will need to be considered.

The process described illustrates the unique combination of sequence and chemistry in silk proteins coupled with the biochemical environment in the gland. These features promote solubility of this hydrophobic protein in water at high concentrations while avoiding premature crystallization into silk II (β -sheets) until fibre spinning. The elucidation of this mechanism also suggests paths forward in polymer design and processing in aqueous environments, even for highly hydrophobic, high-molecular-mass polymers, if the appropriate polymer designs and control of water, and thus the rate of increase in polymer concentration, are considered to achieve desired chain folding and structural transitions. Ultimately these approaches should lead to materials engineering in aqueous systems with desired functional properties, such as for biomaterials and tissue engineering^{30,31}. □

Methods

Protein solutions

Cocoons of *B. mori* silkworm silk were supplied by M. Tsukada (Institute of Sericulture, Tsukuba, Japan). Cocoons were boiled for 30 min in an aqueous solution of $0.02 \text{ M Na}_2\text{CO}_3$, then rinsed thoroughly with water to extract the glue-like sericin proteins. The extracted silk was then dissolved in 9.3 M LiBr solution at room temperature, yielding a $20 \text{ wt}\%$ solution. This solution was dialysed in water using Slide-a-Lyzer dialysis cassettes (Pierce, MWCO 3500) for 48 h. The final concentration of aqueous silk solution was $\sim 8.0 \text{ wt}\%$. Silk blends in water were prepared by adding $5 \text{ wt}\%$ PEO (average molecular weight of $9 \times 10^5 \text{ g mol}^{-1}$) solutions into the silk aqueous solutions. The solutions were stirred for 15 min at room temperature, and then cast on polystyrene Petri dishes and dried for 24 h at room temperature. Under shear conditions, the amount of β -sheet precipitate formed from aqueous fibroin solutions is affected by the solution temperature, fibroin concentration, and stirrer rotation speed¹¹. Therefore, in the present study all blends were gently mixed at low temperature, 7°C , to generate homogeneous solutions before casting. The films were placed in vacuum for another 24 h. For pure fibroin concentrates osmotic stress by dialysis against PEO was used (Fig. 4b).

Sequence analysis

Sequence² was analysed on individual peptide hydrophobicity; the adjacent average method with a window of 10 was used, and iterated 700 times.

AFM

AFM was performed in tapping mode on a Dimension 3100 Nanoscope III with rotated TappingMode etched silicon probes (Nanodevices). Images were obtained in 'phase-mode', and captured at the same time without digital filtering.

Received 22 April; accepted 27 May 2003; doi:10.1038/nature01809.

- Vollrath, F. & Knight, D. P. Liquid crystalline spinning of spider silk. *Nature* **410**, 541–548 (2001).
- Zhou, C. Z. *et al.* Fine organization of *Bombyx mori* fibroin heavy chain gene. *Nucleic Acids Res.* **28**, 2413–2419 (2000).
- Altman, G. H. *et al.* Silk matrix for tissue engineered anterior cruciate ligaments. *Biomaterials* **23**, 4131–4141 (2002).
- Mita, K., Ichimura, S. & James, T. C. Highly repetitive structure and its organization of the silk fibroin gene. *J. Mol. Evol.* **38**, 583–592 (1994).
- Yamada, H., Nakao, H., Takasu, Y. & Tsubouchi, K. Preparation of undegraded native molecular fibroin solution from silkworm cocoons. *Mater. Sci. Eng. C* **14**, 41–46 (2001).
- Sofia, S., McCarthy, M. B., Gronowicz, G. & Kaplan, D. L. Functionalized silk-based biomaterials for bone formation. *J. Biomed. Mater. Res.* **54**, 139–148 (2001).
- Jin, H.-J., Fridrikh, S. V., Rutledge, G. C. & Kaplan, D. L. Electrospinning *Bombyx mori* silk with poly(ethylene oxide). *Biomacromolecules* **3**, 1233–1239 (2002).
- Roseman, M. A. Hydrophilicity of polar amino-acid side-chain is markedly reduced by flanking peptide-bonds. *J. Mol. Biol.* **200**, 513–522 (1988).
- Ochi, A., Hossain, K. S., Magoshi, J. & Nemoto, N. Rheology and dynamic light scattering of silk fibroin solution extracted from the middle division of *Bombyx mori* silkworm. *Biomacromolecules* **3**, 1187–1196 (2002).
- Discher, D. E. & Eisenberg, A. Polymer vesicles. *Science* **297**, 967–973 (2002).
- Malstom, M. & Lindman, B. Self-assembly in aqueous block copolymer solutions. *Macromolecules* **25**, 5440–5445 (1992).
- Kwon, K. W., Park, M. J., Bae, Y. H., Kim, H. D. & Char, K. Gelation behavior of PEO-PLGA-PEO triblock copolymers in water. *Polymer* **43**, 3353–3358 (2002).
- Magoshi, J., Mizuide, M. & Magoshi, Y. Physical properties and structure of silk. VI. Conformational changes in silk fibroin induced by immersion in water at 2 to 130 °C. *J. Polym. Sci.* **17** (Polymer Physics Edition), 515–520 (1979).
- Ishida, M., Asakura, T., Yoko, M. & Saito, H. Solvent- and mechanical-treatment-induced conformational transition of silk fibroins studied by high-resolution solid-state ¹³C NMR spectroscopy. *Macromolecules* **23**, 88–94 (1990).
- Seidel, A. *et al.* Regenerated spider silk: Processing, properties, and structure. *Macromolecules* **33**, 775–780 (2000).
- Valluzzi, R., Szela, S., Avtges, P., Kirschner, D. & Kaplan, D. L. Methionine redox controlled crystallization of biosynthetic silk spidroin. *J. Phys. Chem. B* **103**, 11382–11392 (1999).
- Wilson, D., Valluzzi, R. & Kaplan, D. Conformational transitions in model silk peptides. *Biophys. J.* **78**, 2690–2701 (2001).
- Shen, Y., Johnson, M. A. & Martin, D. C. Microstructural characterization of *Bombyx mori* silk fibers. *Macromolecules* **31**, 8857–8864 (1998).
- Asakura, T., Kuzuhara, A., Tabeta, R. & Saito, H. Conformation characterization of *Bombyx mori* silk fibroin in the solid state by high-frequency ¹³C cross polarization-magic angle spinning NMR, X-ray diffraction, and infra spectroscopy. *Macromolecules* **18**, 1841–1845 (1985).
- Putthanarat, S., Stribeck, N., Fossey, S. A., Eby, R. K. & Adams, W. W. Investigation of the nanofibrils of silk fibers. *Polymer* **41**, 7735–7747 (2000).
- van Beek, J. D., Hess, S., Vollrath, F. & Meier, B. H. The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. *Proc. Natl Acad. Sci. USA* **99**, 10266–10271 (2002).
- Perez-Rigueiro, J., Viney, C., Llorca, J. & Elices, M. Silkworm silk as an engineering material. *J. Appl. Polym. Sci.* **70**, 2439–2447 (1998).
- Poza, P., Perez-Rigueiro, J., Elices, M. & Llorca, J. Fractographic analysis of silkworm and spider silk. *Eng. Fracture Mech.* **69**, 1035–1048 (2002).
- Auvray, X. *et al.* Influence of solvent headgroup interactions on the formation of lyotropic liquid crystal phases of surfactants in water and nonaqueous protic and aprotic solvents. *Langmuir* **8**, 2671–2679 (1992).
- Lele, A. K. *et al.* Deformation induced hydrophobicity: Implications in spider silk formation. *Chem. Eng. Sci.* **56**, 5793–5800 (2001).
- Tanaka, T. *et al.* Phase separation structure in poly(vinyl alcohol) silk fibroin blend films. *Polym. Int.* **45**, 175–184 (1998).
- Knight, D. P. & Vollrath, F. Biological liquid crystal elastomers. *Phil. Trans. R. Soc. Lond. B* **357**, 155–163 (2002).
- Knight, D. P. & Vollrath, F. Liquid crystals and flow elongation in a spider's silk production line. *Proc. R. Soc. Lond. B* **266**, 519–523 (1999).
- Viney, C. Natural silks: Archetypal supramolecular assembly of polymer fibres. *Supramol. Sci.* **4**, 75–81 (1997).
- Minoura, N., Tsukada, M. & Nagura, M. Physico-chemical properties of silk fibroin membrane as a biomaterial. *Biomaterials* **11**, 430–434 (1990).
- Altman, G. H. *et al.* Silk-based biomaterials. *Biomaterials* **24**, 401–416 (2003).

Acknowledgements We thank R. Valluzzi and J. Park for technical input. This work was supported by the NIH, the NSF and the DoD (Air Force).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.L.K. (david.kaplan@tufts.edu).

Cephalopod *Hox* genes and the origin of morphological novelties

Patricia N. Lee^{1,2}, Patrick Callaerts³, Heinz G. de Couet¹ & Mark Q. Martindale²

¹Department of Zoology, University of Hawaii at Manoa, 2538 McCarthy Mall, Honolulu, Hawaii 96822, USA

²Kewalo Marine Laboratory/Pacific Biomedical Research Center, University of Hawaii, 41 Ahui Street, Honolulu, Hawaii 96813, USA

³Department of Biology & Biochemistry, University of Houston, 369 Science and Research Bldg 2, Houston, Texas 77204, USA

Cephalopods are a diverse group of highly derived molluscs, including nautilus, squids, octopuses and cuttlefish. Evolution of the cephalopod body plan from a monoplacophoran-like ancestor¹ entailed the origin of several key morphological innovations contributing to their impressive evolutionary success². Recruitment of regulatory genes³, or even pre-existing regulatory networks⁴, may be a common genetic mechanism for generating new structures. *Hox* genes encode a family of transcriptional regulatory proteins with a highly conserved role in axial patterning in bilaterians⁵; however, examples highlighting the importance of *Hox* gene recruitment for new developmental functions are also known^{6,7}. Here we examined developmental expression patterns for eight out of nine *Hox* genes⁸ in the Hawaiian bobtail squid *Euprymna scolopes*, by whole-mount *in situ* hybridization. Our data show that *Hox* orthologues have been recruited multiple times and in many ways in the origin of new cephalopod structures. The manner in which these genes have been co-opted during cephalopod evolution provides insight to the nature of the molecular mechanisms driving morphological change in the Lophotrochozoa, a clade exhibiting the greatest diversity of body plans in the Metazoa.

Cephalopods share a close affinity with gastropods⁹ (such as snails, limpets, slugs), but have evolved striking modifications of the three defining features of the phylum Mollusca—the shell, mantle and ventral foot (Fig. 1a). Cephalopods have greatly reduced or lost the shell (except *Nautilus*), co-opted the mantle for new functions in locomotion and respiration, and have modified the ventral foot into the prehensile arms of the brachial crown and the funnel tube¹⁰ (Fig. 1a). Accompanying these changes are many sensory and neural innovations, such as well-developed eyes, a vertebrate-like vestibular system, extra-ocular photoreceptors, various types of light organs and complex central (CNS) and peripheral nervous systems¹¹. Here we report the first *Hox* expression patterns in a cephalopod mollusc (Fig. 1b, c). The most conserved and striking feature of the *Hox* gene family is its collinear pattern of expression, in which anterior-class genes are expressed in more rostral domains than posterior-class genes⁵. In the cephalopod *Euprymna scolopes*, however, we have found modifications in these conserved collinear *Hox* expression domains that are correlated with the development of many of the impressive morphological innovations found in this group of molluscs.

Hox expression has been reported in one other mollusc, the abalone *Haliotis asinina*. In this basal gastropod, expression patterns of *Hox* genes suggest that their role in anteroposterior patterning has been conserved in molluscs¹². In the nervous system of the pretorsional veliger, four out of five *Hox* genes examined are collinearly expressed with respect to the rostral–caudal sequence of ganglia development¹². This collinear expression encompasses anterior ganglia corresponding to the cephalopod CNS (for example pleural and pedal) as well as posterior ganglia considered part of the cephalopod peripheral nervous system¹¹ (for example branchial) (Fig. 2a). In the foot rudiment of the trochophore larva, *H. asinina*

Hox2 (*Has-Hox2*) and *Has-Hox3* are expressed in ectodermal domains¹² that correlate with later expression in the pedal ganglia. In the foot of the pre-torsional veliger, *Has-Hox5* is faintly detected in mesodermal regions¹². However, none of the abalone *Hox* orthologues examined is expressed in the developing pedal nerve cords that innervate the gastropod foot.

In *E. scolopes* embryos, six *Hox* genes are expressed in the CNS, which consists of paired cerebral, optic, palliovisceral and pedal ganglia that fuse to form the brain primordium¹¹ (Fig. 2b). *Euprymna scolopes lab* (*Esc-lab*), *Esc-Hox3* and *Esc-Scr* are expressed in regions of the neurogenic placodes of the palliovisceral ganglia^{13,14} (Fig. 2c), and three more posterior genes, *Esc-Antp*, *Esc-Lox4* and *Esc-Post-2*, are expressed in the location of the developing pedal ganglia^{13,14} (Fig. 2d). *Euprymna scolopes Hox* expression within each ganglion does not appear to demonstrate the canonical nested domains characteristic of *Hox* genes⁵, as the anterior expression boundaries are coincident within ganglia. However, the expression pattern along the CNS does not violate collinearity and is consistent with an ancestral role for these genes in axial patterning of the CNS and a presumed clustered genomic arrangement. Two *Hox* orthologues, *Esc-Lox5* and *Esc-Post-1*, seem to have lost this role in *E. scolopes* as neither is expressed in the developing CNS. No *Hox* expression was observed in the anterior-most cerebral ganglia^{13,14}, which may be patterned by cephalic gap genes, or in the optic ganglia, which express a *Pax-6* orthologue¹⁵.

More striking than the expression patterns in the CNS of *E. scolopes* is the observation that *Hox* orthologues are expressed in a number of morphological novelties found only in cephalopods. Specifically, *Hox* genes are expressed in ectodermally derived components of the brachial crown, funnel tube, stellate ganglia, metabranchial vesicles, buccal crown and light organ. However, the anteroposterior expression domains of *Hox* genes in these structures are not those predicted by *Hox* collinearity.

Multiple *Hox* genes are expressed in the brachial crown, a bilaterally symmetrical ring of prehensile arms surrounding the head derived from anterior regions of the molluscan foot¹⁰ (ten in

squids and cuttlefishes, eight in octopuses, and approximately 90 in *Nautilus*). The brachial nervous system is the largest component of the peripheral nervous system and provides neuromuscular control for the complex muscle system of the arms¹¹. Seven of the *Hox* orthologues examined are expressed from the post-gastrula stage in neurogenic regions that give rise to the intrabrachial ganglia of the brachial nervous system^{13,14} (Fig. 3a). As none of the abalone *Hox* orthologues is expressed in the pedal nerve cords of the gastropod foot¹², the *Esc-Hox* expression data provide evidence for the recruitment of at least two, and possibly seven, *Hox* orthologues during the reorganization of the nervous system of the molluscan foot into discrete ganglia of the brachial nervous system. Each arm pair expresses a unique combination of these seven orthologues, raising the possibility of using *Hox* genes as molecular markers for identifying putative arm homologies between cephalopod groups, such as octopuses and squids, or *Nautilus* and coleoids¹⁶. However, this code-like expression is not collinear and evolutionary changes in *Hox* regulation have led to a new and dynamic pattern of expression in the brachial crown (Fig. 3b).

Multiple *Hox* orthologues are also expressed in the stellate ganglia, which are present only in higher cephalopods¹¹ (Fig. 4a). Expansion and co-option of the mantle as a muscular pump for locomotion and respiration were important for the ecological diversification of cephalopods², and the stellate ganglia serve as neural relay centres in the mantle for coordinating the muscle contractions required for these functions. *Esc-lab*, *Esc-Hox3*, *Esc-Lox4* and *Esc-Post-2* are expressed in overlapping domains in the developing stellate ganglia of the peripheral nervous system (Fig. 4b–e). These ganglia possess several neural cell types¹⁷ and the expression data suggest that a subset of *Hox* genes, different from that expressed in the brachial crown, has been coordinately recruited for specification of neuronal cells within these ganglia.

The deployment of multiple *Hox* genes in the brachial crown has not been recruited for specification of all ectodermally derived elements within this structure. The metabranchial vesicles are proposed photoreceptive sensory organs¹⁸ located in the brachial crown

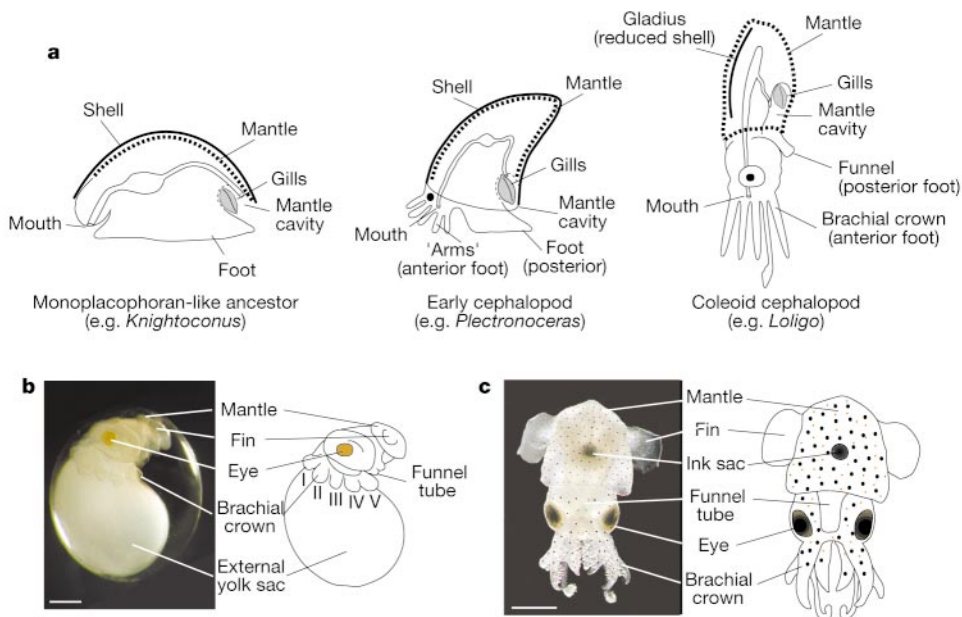


Figure 1 Cephalopod evolution and body plan. **a**, Proposed scheme³⁰ of cephalopod evolution from a monoplacophoran-like ancestor, illustrating shell internalization/reduction, mantle expansion and co-option for locomotion/respiration, and foot modifications forming the brachial crown and funnel tube¹⁰ (anterior, left; dorsal, top). **b**, *Euprymna scolopes* embryo, stage 21–22, and diagram illustrating morphological

features of the cephalopod body plan. Brachial crown arm pairs are numbered I–V. Arm IV is the specialized feeding tentacle (anterior, left; dorsal, top). Scale bar, 0.5 mm. **c**, Posterior view of *E. scolopes* hatchling (dorsal, top), and diagram showing the characteristic adult cephalopod body plan. Scale bar, 1 mm.

between arms I and II (superior anterior vesicles) and arms IV and V (inferior anterior vesicles)¹⁹. In contrast to *Hox* expression in the brachial crown and stellate ganglia, *Esc-Hox3* is the only orthologue we examined that is expressed in the superior and inferior anterior vesicles (Fig. 5a, b), but is not expressed elsewhere in the peripheral nervous system of the brachial crown (Fig. 3a). Development of the metabranchial vesicles, therefore, does not involve a combination of *Hox* orthologues as seen for arms of the brachial crown, although

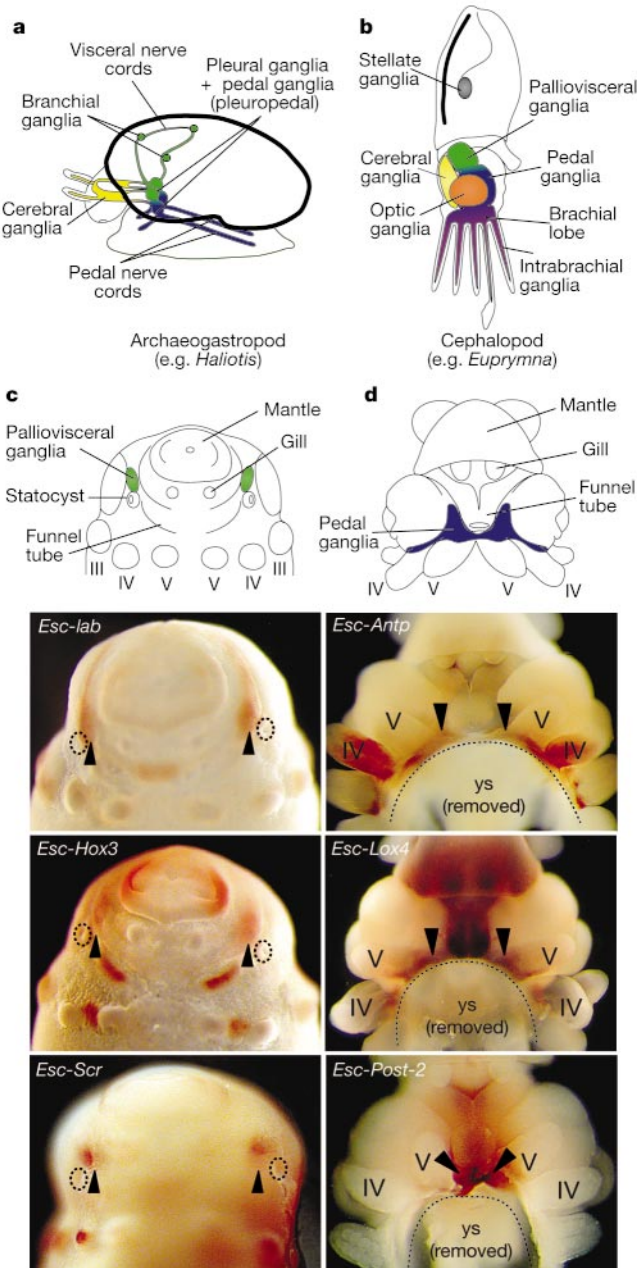


Figure 2 Evolution of the cephalopod brain and *Hox* expression in the CNS. **a, b**, Comparison of nervous systems in archaeogastropods (**a**), similar to the archetypal mollusc, with two pairs of nerve cords extending from paired ganglia, and cephalopods (**b**), illustrating centralization/fusion of ganglia, addition of optic ganglia, and elaboration of the brachial lobe from the pedal ganglia (anterior, left; dorsal, top). **c**, Top, stage 19–20 embryo (posterior) indicating location of the paired palliovisceral ganglia. Below, *Esc-lab*, *Esc-Hox3* and *Esc-Scr* expression in *E. scolopes* embryos (arrowheads; statocysts are circled). **d**, Top, stage 24–25 embryo (posterior) showing the paired pedal ganglia. Below, *Esc-Antp*, *Esc-Lox4* and *Esc-Post-2* expression in *E. scolopes* embryos (arrowheads). ys, yolk sac.

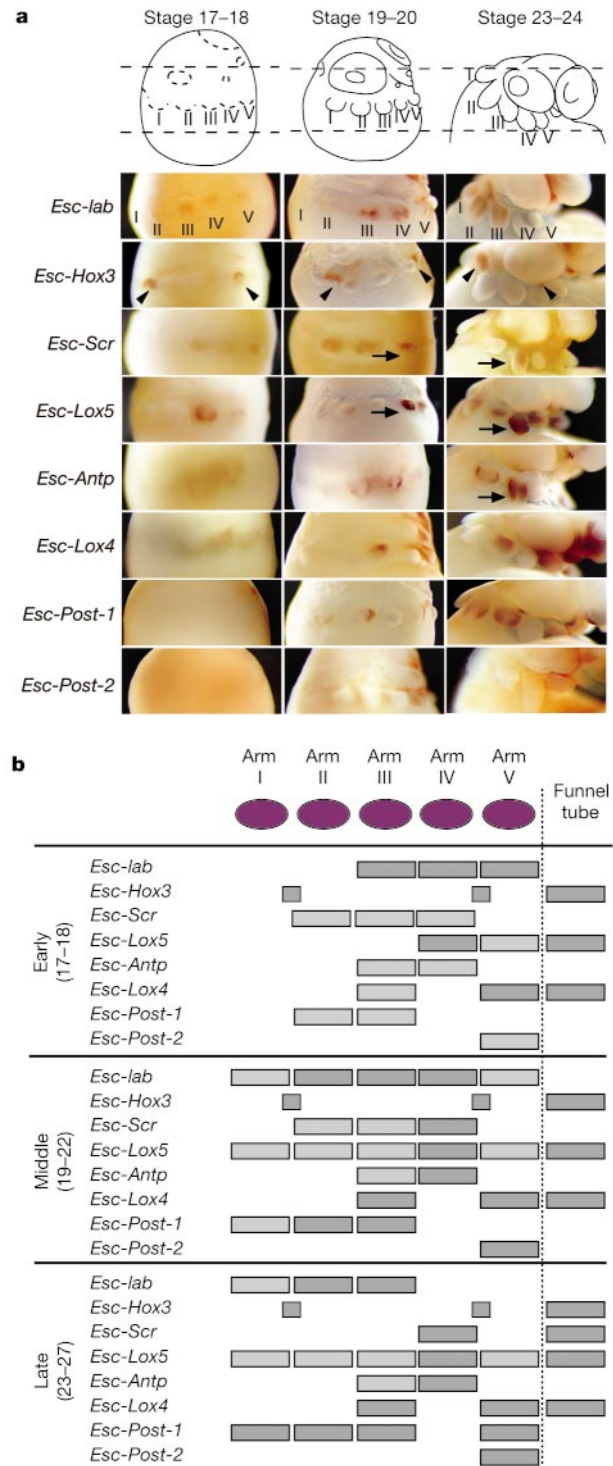


Figure 3 *Hox* expression in the developing brachial crown. **a**, Top, diagrams of embryos for each stage; dotted lines indicate embryo regions shown below; arm pairs are numbered. Bottom, *Hox in situ* hybridization on whole-mount embryos (anterior, left; dorsal, top). *Esc-Hox3* is expressed in metabranchial vesicles (arrowheads), not in the arms. *Esc-Scr*, *Esc-Lox5* and *Esc-Antp* are strongly expressed in the feeding tentacle (arrows). Most orthologues are expressed weakly at stage 17–18. **b**, Summary of the dynamic temporal and spatial pattern of *Hox* expression in the brachial crown and funnel tube during development (relative expression levels indicated by shading intensity).

the vesicles are located in the same axial region as specific arm pairs. This suggests that in the brachial crown these *Hox* genes are not specifying anteroposterior position, but instead have been recruited independently for development of these different structures.

Esc-Hox3 expression in the metabranchial vesicles is not the only example of individual gene recruitment for the development of new cephalopod structures. The buccal crown of squids, also derived from the anterior of the molluscan foot, is a ring of seven small appendages (lappets) within the brachial crown and surrounding the mouth¹⁶ (Fig. 5e). In contrast to the code-like *Hox* expression in the brachial crown, *Esc-Post-1* is the only orthologue expressed from early organogenesis in regions of the developing buccal lappets (Fig. 5c, d, f). Additionally, *Esc-Post-1* is the only orthologue examined that is expressed in the developing light organ (Fig. 5c, d, h), a complex bilobed structure located posteriorly in the mantle cavity (Fig. 5g). The light organ houses a symbiotic bioluminescent bacterium, which may provide counter-illumination used in predator avoidance²⁰. *Esc-Post-1* expression in the brachial crown, buccal crown and light organ illustrates that the development of new structures in different positions along the anteroposterior axis may involve the same *Hox* gene. Our expression patterns for *Esc-Post-1*, as well as other examples of individual^{7,21,22} and coordinated

recruitment²³ of posterior *Hox* orthologues in other bilaterians, suggest that posterior-class genes may be more readily co-opted for new developmental programmes.

Hox expression patterns in three innovations derived from the molluscan foot highlight the various ways in which *Hox* genes have been recruited in the cephalopod *E. scolopes*. For example, both the brachial crown and buccal crown are derived from the anterior foot and have a nearly one-to-one spatial correspondence between the brachial and buccal arms (Fig. 5e). However, the arms of the brachial crown express a non-collinear and dynamic combination

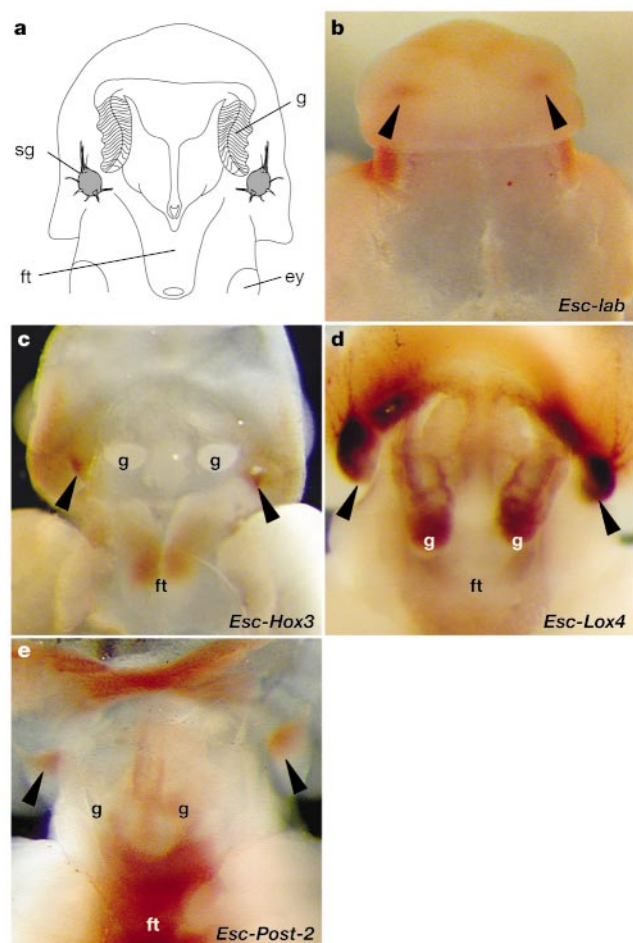


Figure 4 *Hox* expression in the stellate ganglia. **a**, Diagram of a posterior view of the mantle cavity and location of the stellate ganglia. **b**, *Esc-lab* expression (arrowheads) in the stellate ganglia, on the inner mantle surface near the fin bases (anterior view). **c**, *Esc-Hox3* expression (arrowheads) in the stellate ganglia (posterior view, mantle dissected along midline). **d**, *Esc-Lox4* expression in the stellate ganglia and nerves (posterior view). **e**, *Esc-Post-2* expression in the stellate ganglia (posterior view). ey, eye; ft, funnel tube; g, gill; sg, stellate ganglion.

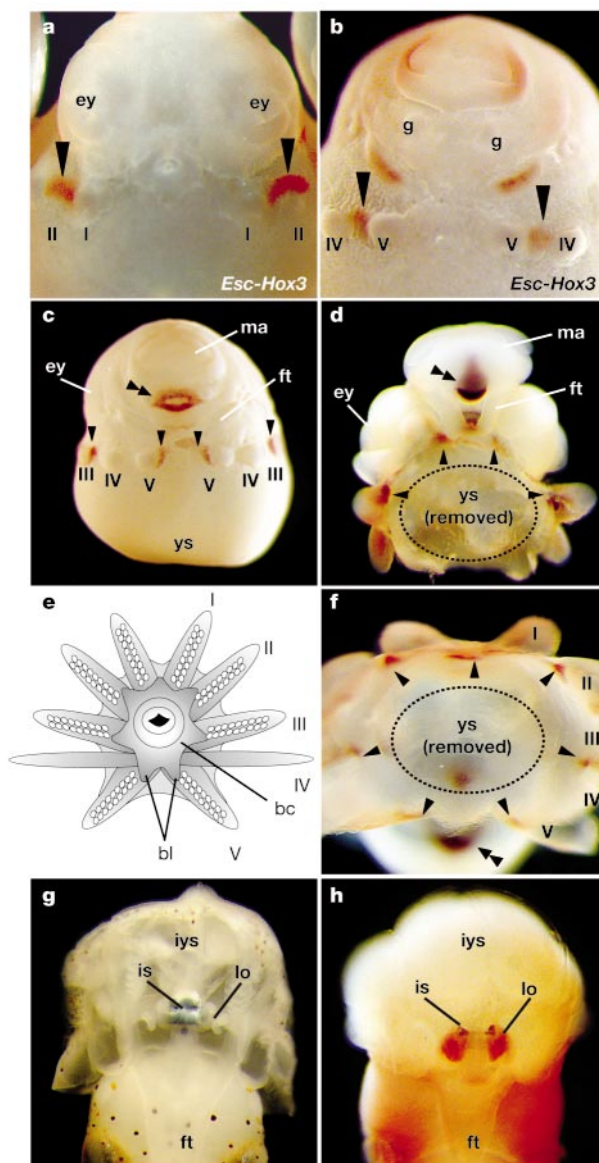


Figure 5 Individual *Hox* expression in metabranchial vesicles, buccal crown and light organ. **a**, **b**, *Esc-Hox3* expression in superior anterior vesicles (**a**) and inferior anterior vesicles (**b**) (arrowheads; stage 19–20; **a**, anterior; **b**, posterior). **c**, *Esc-Post-1* expression in the lappets (arrowheads) and light organ primordium (double arrowhead; stage 19–20; posterior). **d**, *Esc-Post-1* expression in the lappets (arrowheads) and light organ (double arrowhead) (stage 23–24; posterior). **e**, Diagram of a ventral view through the brachial crown illustrating positions of the lappets. **f**, Ventral view through the brachial crown showing *Esc-Post-1* expression in the lappets (arrowheads) and light organ (double arrowheads; stage 25–26). **g**, Posterior view of hatchling showing light organ location in the mantle cavity. **h**, Posterior view (mantle removed) of *Esc-Post-1* in the light organ (stage 29–30). bc, buccal crown; bl, buccal lappet; ey, eye; ft, funnel tube; is, ink sac; iys, internal yolk sac; lo, light organ; ma, mantle; ys, external yolk sac.

of anterior-, central- and posterior-class *Hox* orthologues, whereas the lappets of the buccal crown express a single posterior *Hox* gene. The funnel tube, which is derived from the posterior of the molluscan foot, expresses a subset of four *Hox* orthologues, including several more anterior- and central-class genes (for example, *Esc-Hox3* and *Esc-Scr*) (Fig. 3b). The specific patterns of *Hox* expression in each of these structures suggest that the acquisition of additional regulatory elements, functioning independently of collinearity, may be associated with the distinct evolutionary trajectories for three morphologies derived from a common ancestral structure.

Hox genes have been correlated with key events in body plan evolution in ecdysozoans²⁴ and deuterostomes²⁵. Here we have shown that specific *Hox* expression is developmentally correlated with a number of morphological innovations in a lophotrochozoan. Expression of multiple *Hox* genes in the brachial crown, funnel tube and stellate ganglia in cephalopods is the first reported non-vertebrate example of coordinated *Hox* expression in evolutionary novelties, although it does not reflect the expected collinear pattern. The *Hox* expression patterns suggest that in this cephalopod, the co-option of subsets of the *Hox* complex is not necessarily constrained by the regulatory mechanisms of collinearity seen with multiple *Hox* recruitment in vertebrate innovations (for example limbs, genitalia). In fact, our expression data indicate that *Hox* genes can potentially be co-opted to function in diverse developmental contexts, not all of which are related to patterning the anteroposterior axis. Although this is one of the first molecular investigations of body plan evolution in the Mollusca, our data, and previously reported *Hox*¹² and *engrailed*^{26–28} expression patterns in other molluscs, suggest that the morphological plasticity that is so striking in this group of lophotrochozoans may be due to relaxation of regulatory constraints on the recruitment of various regulatory genes involved in morphological patterning of bilaterian animals. □

Methods

Collection of *E. scolopes* embryos

Embryos were collected, prepared, fixed and staged as previously described¹⁵.

Riboprobe synthesis

Euprymna scolopes *Hox* complementary DNA clones were generated by rapid amplification of cloned ends (RACE)—PCR as previously described⁸. Digoxigenin-labelled anti-sense and sense riboprobes were transcribed *in vitro* from linearized 5' RACE clones (MEGAscript High Yield Transcription Kit; Ambion) for *Esc-lab*, *Esc-Scr*, *Esc-Lox5*, *Esc-Antp*, *Esc-Lox4*, *Esc-Post-1* and *Esc-Post-2*. For *Esc-Hox3*, a longer cDNA from merged 5' and 3' RACE sequences was generated by PCR and linearized for *in vitro* transcription.

Whole-mount *in situ* hybridization

The whole-mount *in situ* hybridization protocol used here is modified after that of ref. 29. Fixed embryos were treated with 0.01 mg ml⁻¹ proteinase K (GibcoBRL) in PTw (1 × PBS, 0.1% Tween-20) for 8–20 min at room temperature, and then post-fixed in 3.7% formaldehyde in PTw for 1 h at room temperature. Embryos were washed five times for 5 min each in PTw at room temperature, then heated for 15 min at 80 °C in PTw to inactivate endogenous alkaline phosphatases. Embryos were incubated in pre-hybridization buffer (50% formamide; × 5 SSC, pH 4.5; 50 µg ml⁻¹ heparin, 0.1% Tween-20; 1% SDS; 100 µg ml⁻¹ denatured single-strand salmon sperm DNA) for 10 min at room temperature, then at 65 °C overnight in fresh pre-hybridization buffer. Embryos were hybridized with sense or anti-sense riboprobe (0.1–0.5 ng µl⁻¹) at 65 °C for 18–24 h.

After hybridization, embryos were washed for 20 min in pre-hybridization buffer at 65 °C, then for 15 min each in 75% pre-hybridization buffer and 25% × 2 SSC; 50% pre-hybridization buffer and 50% × 2 SSC; 25% pre-hybridization buffer and 75% × 2 SSC; and 100% × 2 SSC. Two additional 30-min washes in × 0.05 SSC were performed at 65 °C. Embryos were then washed for 10 min each at room temperature in 75% × 0.05 SSC and 25% PTw; 50% × 0.05 SSC and 50% PTw; 25% × 0.05 SSC and 75% PTw; and 100% PTw. Embryos were blocked by washing five times for 5 min each in PBT (× 1 PBS; 0.1% BSA; 0.2% Triton X-100) at room temperature, and then in 1% blocking buffer (Boehringer-Mannheim blocking powder in 100 mM maleic acid, 150 mM NaCl, pH 7.5) for 2 h at room temperature. Embryos were incubated in anti-digoxigenin-AP antibody (1:2,500) (Boehringer-Mannheim) at 4 °C overnight, then washed eight times for 10 min each in PBT at room temperature. Embryos were incubated in AP buffer (100 mM NaCl, 50 mM MgCl₂, 100 mM Tris, pH 9.5, 0.1% Tween-20) and riboprobes were detected colourimetrically (NBT/BCIP). Whole-mount embryos were observed under a Wild M8 stereomicroscope, and digital photomicrographs were taken with a Nikon Cool Pix 990 digital camera (3.34 megapixel).

Sequences of cDNA fragments

Sequences of cDNA fragments used in this study have been deposited in GenBank under the following accession codes: AY330184 (5' end of *Esc-lab*), AY30185 (*Esc-Hox3*), AY330186 (5' end of *Esc-Scr*), AY330187 (5' end of *Esc-Lox5*), AY330188 (5' end of *Esc-Antp*), AY330189 (5' end of *Esc-Lox4*), AY330190 (5' end of *Esc-Post-1*), AY330191 (5' end of *Esc-Post-2*).

Received 17 January; accepted 1 July 2003; doi:10.1038/nature01872.

- Runnegar, B. & Pojeta, J. J. Molluscan phylogeny: the paleontological viewpoint. *Science* **186**, 311–317 (1974).
- House, M. R. in *Cephalopods—Present and Past* (eds Wiedmann, J. & Kullmann, J.) 1–16 (Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 1988).
- Lowe, C. J. & Wray, G. A. Radical alterations in the roles of homeobox genes during echinoderm evolution. *Nature* **389**, 718–722 (1997).
- Keys, D. N. et al. Recruitment of a *hedgehog* regulatory circuit in butterfly eyespot evolution. *Science* **283**, 532–534 (1999).
- Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Malden, 2001).
- Telford, M. J. Evidence for the derivation of the *Drosophila fushi tarazu* gene from a *Hox* gene orthologous to lophotrochozoan. *Lox5. Curr. Biol.* **10**, 349–352 (2000).
- Godwin, A. R. & Capecchi, M. R. *Hox13* mutant mice lack external hair. *Genes Dev.* **12**, 11–20 (1998).
- Callaerts, P. et al. *HOX* genes in the sepiolid squid *Euprymna scolopes*: implications for the evolution of complex body plans. *Proc. Natl Acad. Sci. USA* **99**, 2088–2093 (2002).
- Brusca, R. C. & Brusca, G. J. *Invertebrates* (Massachusetts, Sinauer, Sunderland, 2003).
- Boletzky, S. v. in *Cephalopods—Present and Past* (ed. Wiedmann, J. K., J.) 167–179 (Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 1988).
- Budelmann, B. U. in *The Nervous Systems of Invertebrates: An Evolutionary and Comparative Approach* (eds Breidbach, O. & Kutsch, W.) 115–138 (Birkhauser, Basel, 1995).
- Hinman, V. F., O'Brien, E. K., Richards, G. & Degan, B. M. Expression of anterior *Hox* genes during larval development of the gastropod *Halotis asinina*. *Evol. Dev.* (in the press).
- Shigeno, S., Kidokoro, H., Tsuchiya, K., Segawa, S. & Yamamoto, M. Development of the brain in the oegopsid squid, *Todarodes pacificus*: an atlas up to the hatching stage. *Zool. Sci.* **18**, 527–541 (2001).
- Shigeno, S., Tsuchiya, K. & Segawa, S. Embryonic and paralarval development of the central nervous system of the loliginid squid *Sepioteuthis lessoniana*. *J. Comp. Neurol.* **437**, 449–475 (2001).
- Hartmann, B. et al. Pax6 in the sepiolid squid, *Euprymna scolopes*: evidence for a role in eye, sensory organ and brain development. *Mech. Dev.* **120**, 177–183 (2003).
- Boletzky, S. v. *McGraw-Hill Yearbook of Science & Technology* 73–76 (McGraw-Hill, New York, 1994).
- Young, J. Z. The organization of a cephalopod ganglion. *Phil. Trans. R. Soc. Lond. B* **263**, 409–429 (1972).
- Boletzky, S. v., Frösch, D. & Mangold, K. Développement de vésicules associées au complexe brachial chez les Céphalopodes. *C.R. Acad. Sci. (Paris) D* **270**, 2182–2184 (1995).
- Sundermann, G. Development and hatching state of ectodermal vesicle-organs in the head of *Sepia officinalis*, *Loligo vulgaris*, and *Loligo forbesi* (Cephalopoda, Decabrachia). *Zoomorphology* **109**, 343–352 (1990).
- McFall-Ngai, M. & Montgomery, M. The anatomy and morphology of the adult bacterial light organ of *Euprymna scolopes* Berry (Cephalopoda: Sepiolidae). *Biol. Bull.* **179**, 332–339 (1990).
- Nogi, T. & Watanabe, K. Position-specific and non-colinear expression of the planarian posterior (Abdominal-B-like) gene. *Dev. Growth Diff.* **43**, 177–184 (2001).
- Kulakov, M. A., Kostyuchenko, R. P., Andreeva, T. F. & Dondua, A. K. The Abdominal-B-like gene expression during larval development of *Nereis virens* (polychaeta). *Mech. Dev.* **115**, 177–179 (2002).
- Zákány, J. & Duboule, D. *Hox* genes in digit development and evolution. *Cell Tissue Res.* **296**, 19–25 (1998).
- Averof, M. & Patel, N. H. Crustacean appendage evolution associated with changes in *Hox* gene expression. *Nature* **388**, 682–686 (1997).
- Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. *Hox* genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1995).
- Moshel, S. M., Levine, M. & Collier, J. R. Shell differentiation and *engrailed* expression in the *Ilyanassa* embryo. *Dev. Genes Evol.* **208**, 135–141 (1998).
- Nederbragt, A. J., van Loon, A. E. & Dictus, W. J. A. G. Expression of *Patella vulgata* orthologs of *engrailed* and *dpp-BMP2/4* in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. *Dev. Biol.* **246**, 341–355 (2002).
- Wanninger, A. & Haszprunar, G. The expression of an *engrailed* protein during embryonic shell formation in the tusk-shell, *Antalis entalis* (Mollusca, Scaphopoda). *Evol. Dev.* **3**, 312–321 (2001).
- Seaver, E. C., Paulson, D., Irvine, S. Q. & Martindale, M. Q. The spatial and temporal expression of *Ch-en*, the *engrailed* gene in the polychaete *Chaetopterus*, does not support a role in body axis segmentation. *Dev. Biol.* **236**, 195–209 (2001).
- Holland, C. H. The nautiloid cephalopods: a strange success. *J. Geol. Soc. Lond.* **144**, 1–15 (1987).

Acknowledgements We thank D. Paulson for assistance with *in situ* hybridization protocols; E. Seaver, A. Wikramanayake, L. Nederbragt, D. Matus, K. Pang and L. Kaneshige for critical reading of the manuscript; B. Degan for sharing unpublished data; S. v. Boletzky for sending us reprints of many cephalopod developmental studies; and M. McFall-Ngai for providing additional embryos. This study was supported by grants from the National Science Foundation to M.Q.M., H.G.d.C. and P.C. Kewalo Marine Laboratory/Pacific Biomedical Research Center, University of Hawaii is the present address of P.N.L.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.G.d.C. (couet@hawaii.edu).

Suppression of CED-3-independent apoptosis by mitochondrial β NAC in *Caenorhabditis elegans*

Tim A. Bloss*, Eric S. Witzke* & Joel H. Rothman

Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara, California 93106, USA

* These authors contributed equally to this work

To ensure cell survival, it is essential that the ubiquitous pro-apoptotic machinery is kept quiescent. As death is irreversible, cells must continually integrate developmental information with regulatory inputs to control the switch between repressing and activating apoptosis. Inappropriate activation or suppression of apoptosis can lead to degenerative pathologies¹ or tumorigenesis², respectively. Here we report that *Caenorhabditis elegans* inhibitor of cell death-1 (ICD-1) is necessary and sufficient to prevent apoptosis. Loss of ICD-1 leads to inappropriate apoptosis in developing and differentiated cells in various tissues. Although this apoptosis requires CED-4, it occurs independently of CED-3—the caspase essential for developmental apoptosis³—showing that these core pro-apoptotic proteins have separable roles. Overexpressing ICD-1 inhibits the apoptosis of cells that are normally programmed to die. ICD-1 is the β -subunit of the nascent polypeptide-associated complex (β NAC) and contains a putative caspase-cleavage site and caspase recruitment domain. It localizes primarily to mitochondria, underscoring the role of mitochondria in coordinating apoptosis⁴. Human β NAC is a caspase substrate that is rapidly eliminated in dying cells^{5,6}, suggesting that ICD-1 apoptosis-suppressing activity may be inactivated by caspases.

Genetic studies in *C. elegans* have shown that three conserved components of the core apoptotic regulatory pathway, CED-3, CED-4 (Apaf1 in mammals) and CED-9, control the switch between survival and death (reviewed in ref. 3). CED-9, a member of the Bcl-2 family, is the pivotal suppressor that prevents CED-4/Apaf1 from promoting activation of the pro-apoptotic CED-3 caspase³. In the absence of either CED-3 or CED-4, almost all apoptosis is abrogated⁷, and the pro-apoptotic requirements of these two core regulators have been thought to be inseparable. The organelle that orchestrates much of the control of apoptosis is the mitochondrion, where essential pro- and anti-apoptotic factors including CED-4 and CED-9 are localized in living cells³. Here we have identified ICD-1 as a suppressor of apoptosis that may regulate the core apoptotic programme in mitochondria.

While analysing candidate regulators of early embryogenesis, we found that debilitating the function of *icd-1* by RNA-mediated interference (RNAi) results in a marked increase in cell death in *C. elegans*. Eliminating all immunologically detectable ICD-1 results in early embryonic arrest (at >30 cells, well before apoptosis normally occurs), whereas reducing, but not eliminating, ICD-1 leads to an increase in cell death in later embryonic and larval stages (Fig. 1a, b, and data not shown). An increase in refractile cell corpses⁸ is first observed at the 'comma' stage of development (~300 min post-fertilization (m.p.f.)), shortly after developmental apoptosis begins. At this stage (Fig. 1c), *icd-1(RNAi)* embryos contain more than half again as many corpses (10.6 ± 5.6) as do control embryos (6.6 ± 2.3 ; Wilcoxon rank sum, $P = 3 \times 10^{-4}$). As embryogenesis proceeds, late embryos often arrest and become filled with numerous cell corpses (Fig. 1b). Although *icd-1(RNAi)* embryos often die, many hatch into larvae that are heavily laden with such corpses, particularly in regions of the nervous system. A

few of the observed corpses persist for abnormally long periods; however, most are engulfed over a normal time course, implying that the large increase in corpse number is the result of increased cell death rather than diminished engulfment.

On the basis of their morphology, the refractile structures that we observe (Fig. 1b) seem to be apoptotic cell corpses. These structures, including some that are misshapen or abnormally small (Fig. 1b, white arrowheads), stain with the dye SYTO 12 (data not shown), indicating that they are dead cells containing remnants of nuclei. Numerous electron-dense structures are visible by electron microscopy, particularly in the ventral nerve cord and nerve ring; these structures correlate with cell corpses seen in the same regions

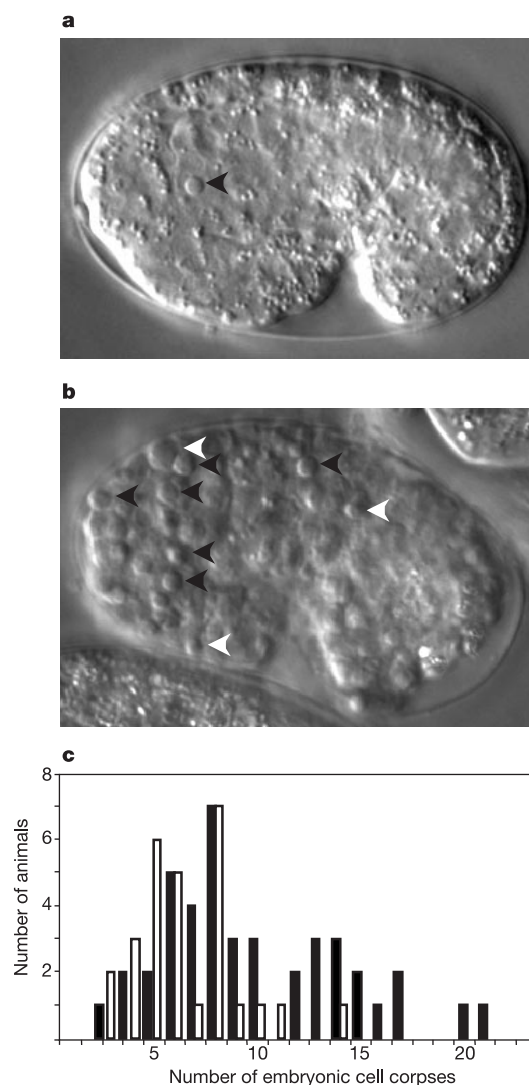


Figure 1 Reduction of ICD-1 activity results in increased apoptosis. **a**, Nomarski image of a wild-type (N2) embryo with one cell corpse showing the characteristic round, refractile appearance⁸ (arrowhead). **b**, Terminal *icd-1(RNAi)* embryo with more than 40 cell corpses. Although most of these have the characteristic apoptotic morphology (black arrowheads), some appear abnormal in size and shape (white arrowheads). Such abnormal cell corpses are particularly conspicuous in later embryos and larvae (see Fig. 5). **c**, Quantification of cell corpses in *icd-1(RNAi)* embryos. Developing embryos were scored for cell corpse number at the comma stage of development (~300 m.p.f.). Embryos from both *icd-1(RNAi)* (filled bars) and *unc-22(RNAi)* (open bars) control populations were scored. *unc-22(RNAi)* results in an uncoordinated phenotype but has no conspicuous effect on apoptosis. With the exception of the increased apoptosis in the former, *icd-1(RNAi)* and *unc-22(RNAi)* embryos are morphologically indistinguishable.

by Nomarski optics (data not shown). The corpses do not have the vacuolated appearance of necrotic⁹ or pathological¹⁰ cell deaths.

The increase in apoptosis in *icd-1(RNAi)* embryos and larvae is correlated with a loss of various cell types. Neurons seem to be particularly sensitive to *icd-1(RNAi)*, because large numbers of cells are absent in the ventral nerve cord and nerve ring, as assessed by a

neuron-specific green fluorescent protein (GFP) marker (Fig. 2a, b). In a few extreme cases, most nerve ring neurons are absent (Fig. 2b), and such animals often show voids in the region normally occupied by the nerve ring (data not shown).

We confirmed that the lack of GFP corresponds to an absence of cells and not simply to decreased marker expression by quantifying nuclei stained with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) anterior to and including the retro-vesicular ganglion (the most posterior region of the nerve ring). In the animals quantified, as many as a third of the normal number of nuclei in the region were missing, and affected animals contained 240 ± 19 anterior nuclei, as compared with 275 ± 24 in the wild-type population ($P = 3 \times 10^{-3}$). The absence of neurons is due to ectopic cell death rather than to a failure to produce neurons during development, because profoundly affected animals contain roughly normal numbers of neurons earlier in development before the widespread cell death is observed. Consistent with a loss of neurons, many affected animals are severely uncoordinated.

Neurons and structural cells were also found to be missing in the tail of *icd-1(RNAi)* adult males. The male tail is a fan-like structure containing 18 sensory projections, or rays¹¹. Each ray consists of three cells, two neurons and a structural cell, all of which share a common progenitor (Fig. 2c). A missing ray can indicate the death of a progenitor cell, whereas a deformed ray implies the loss of one or more of the descendant cells (Fig. 2c). We found that *icd-1(RNAi)* males were missing nearly half of the number of normal rays (10.4 ± 6.0 versus 17.1 ± 1.0 normal rays in wild-type), and lacked nearly a third of the number of total rays (12.3 ± 4.7 versus 17.2 ± 0.8 normal and deformed rays in wild-type; $P = 1 \times 10^{-4}$; Fig. 2d–f). The number of missing and deformed rays is comparable to that seen in mutants defective for CED-9 (ref. 12). This loss of rays is apparently the result of inappropriate apoptosis rather than a cell lineage alteration, on the basis of its dependence on the apoptotic regulatory machinery.

Apoptosis in *C. elegans* occurs frequently in neuronal cell lineages. We therefore examined whether ICD-1 is also required to prevent apoptosis in lineages and tissue types in which cell death does not occur. Indeed, we found that intestinal cells, which arise

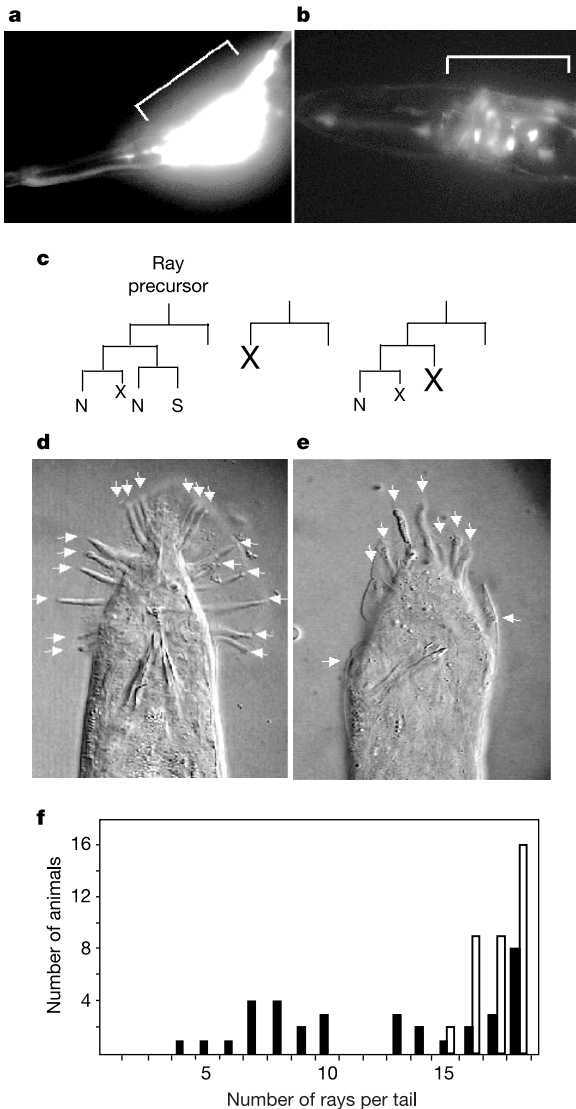


Figure 2 Loss of neurons in *icd-1(RNAi)* animals. **a**, Head region of a *him-8;edls20* larva in which all neurons are marked by GFP. Anterior is to the left. The roughly 200 neurons that constitute the nerve ring¹³ (bracket) are identified by fluorescence. The intense signal generated by these cells masks the individual neurons in the ring. **b**, Nerve ring of a severely affected *icd-1(RNAi);him-8;edls20* larva. The loss of neurons in this region is reflected by a greatly diminished GFP signal, which allows individual surviving neurons to be seen. **c**, Cell lineage of male tail rays. Left, lineage of a normal ray, comprising two neurons (N) and a structural cell (S)¹¹. The X signifies a normal apoptotic death that occurs during ray development. Middle, example of an inappropriate apoptotic death (enlarged X) that would eliminate all cells of a ray, resulting in its complete loss. Right, example of an inappropriate apoptotic death that would result in a deformed ray. The loss of any cell of the ray (enlarged X) would result in a ray with an abnormal morphology. **d**, Tail fan of a wild-type male containing 18 rays (arrows). Anterior is down. **e**, Tail fan of an *icd-1(RNAi);him-8;edls20* male with nine missing rays. Of the nine remaining rays, most are deformed. **f**, Quantification of rays in an *icd-1(RNAi)* male. *icd-1(RNAi);him-8;edls20* males (filled bars) were scored for the number of wild-type and deformed rays. *unc-22(RNAi);him-8;edls20* males were scored as controls (open bars).

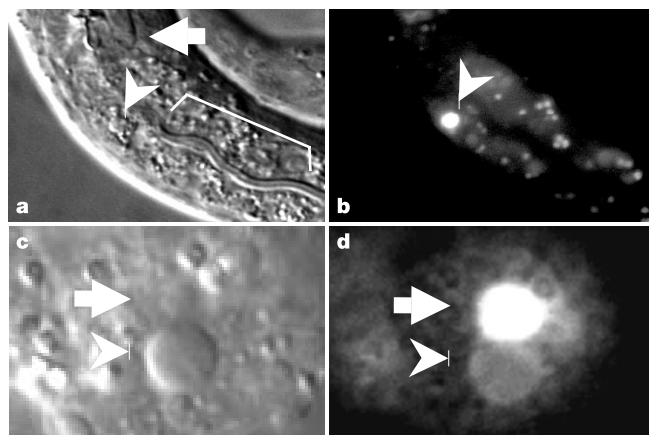


Figure 3 Reduction of ICD-1 activity leads to apoptotic death of intestinal cells.

a, Nomarski image of a cell corpse (arrowhead) in the gut of an *icd-1(RNAi)* larva. This corpse was observed posterior to the pharynx (arrow) and proximal to the gut lumen (bracket). **b**, SYTO 12 staining of the corpse in **a**. **c**, Nomarski image of a cell corpse (arrowhead) in the gut region of a severely affected *icd-1(RNAi);elt-2::GFP* embryo. It is larger than a normal cell corpse, presumably because gut cells, which do not normally undergo apoptosis, are the largest cells in the animal. **d**, Fluorescence image of the embryo shown in **c**. ELT-2 is expressed exclusively in differentiating intestinal cells²⁸. The cell corpse (arrowhead) retains small amounts of ELT-2-GFP, whereas the nucleus of the adjacent surviving intestinal cell (arrow) expresses large quantities of the marker.

from a lineage in which apoptosis does not normally occur¹³, often die in *icd-1(RNAi)* mutants (Fig. 3). On the basis of nuclear staining with propidium iodide, about 20% ($n = 53$) of *icd-1(RNAi)* L1 larvae contained fewer than the normal 20 gut cells (16.5 ± 2.4), and as few as 10 gut cells were observed. In addition, we observed SYTO-12-reactive cells and cells with the characteristic apoptotic morphology in the intestines of *icd-1(RNAi)* larvae and embryos (Fig. 3a, b) and found that the core apoptotic machinery is essential for this death of intestinal cells. Compared with other cell types, gut cells are relatively insensitive to apoptosis: a minority of larvae and only the most severely affected embryos show cell death in the intestine, on the basis of missing gut cells. This resistance to apoptosis may reflect the high concentrations of ICD-1 that are expressed in intestinal cells, which may make them less susceptible to RNAi-induced decreases in ICD-1.

Additional observations show that differentiated cells can undergo apoptotic death in *icd-1(RNAi)* animals. For example, apoptosis in these mutants can occur in the nerve ring many hours after neurons have differentiated. In addition, a marker of the differentiating intestine can be detected in gut cells undergoing apoptosis (*elt-2::gfp*; Fig. 3c, d). The finding that cells possessing a gut identity undergo apoptosis also implies that the inappropriate apoptosis in *icd-1(RNAi)* mutants is a result of derepressing the apoptotic programme *per se*, rather than an alteration in cell lineage or identity. The death of differentiated cells may explain why some corpses are abnormally small and irregular in shape (Fig. 1b), in contrast to those that arise from undifferentiated spherical cells in normal development. These abnormal corpses may arise from the death of cells that have already differentiated and undergone morphological changes.

As our results suggested that ICD-1 is required to suppress apoptosis in normally surviving cells, we examined whether ICD-1 is sufficient to prevent apoptosis of cells normally fated to die by assessing the effect of overexpressing it from a ubiquitous heat-shock promoter. Embryos overexpressing ICD-1 contained a third fewer corpses (16.4 ± 3.8) at the comma stage than did control embryos (23.6 ± 4.5) carrying the same corpse engulfment-defective mutation ($P = 2 \times 10^{-7}$; Fig. 4). Concomitant with the diminished corpse number was an increase in the number of surviving cells in resulting larvae, as assessed by counting live nuclei in the anterior pharynx (data not shown). The range (0–11) and average number (3.8 ± 2.7) of extra surviving cells in the pharynxes were comparable to those seen in animals expressing

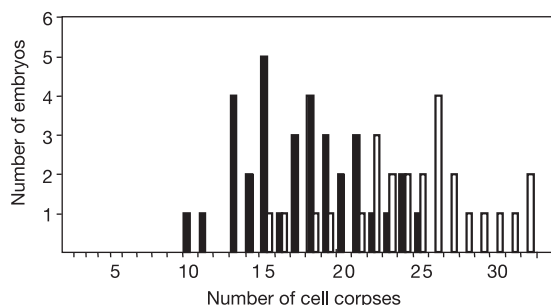


Figure 4 Overexpression of ICD-1 suppresses developmentally programmed apoptosis. Apoptosis was quantified in *ced-1(e1735)* embryos overexpressing ICD-1. Embryos containing an integrated construct expressing ICD-1 from the *hsp16-2* and *hsp16-41* heat-shock promoters were scored for cell corpses at comma stage (filled bars). The *ced-1(e1735)* mutation results in a cell-corpse engulfment defect that causes corpses to persist⁸. Resulting larvae from these experiments often contain extra cells in their pharynxes, but otherwise appear normal. As a control, *ced-1(e1735)* embryos without a heat-shock construct were heat shocked and scored for number of cell corpses (open bars). Suppression of apoptosis by overexpression of ICD-1 was confirmed in two independent transgenic lines.

baculovirus p35, a strong caspase inhibitor¹⁴. These surviving cells were similar in location and morphology to those in apoptosis-deficient *ced-3* mutants⁷, indicating that they arose from suppression of CED-3-dependent developmental apoptosis.

The excessive cell death observed in *icd-1(RNAi)* animals might be triggered by inappropriate derepression of the normal apoptotic programme or instead by an alternative mechanism. We therefore

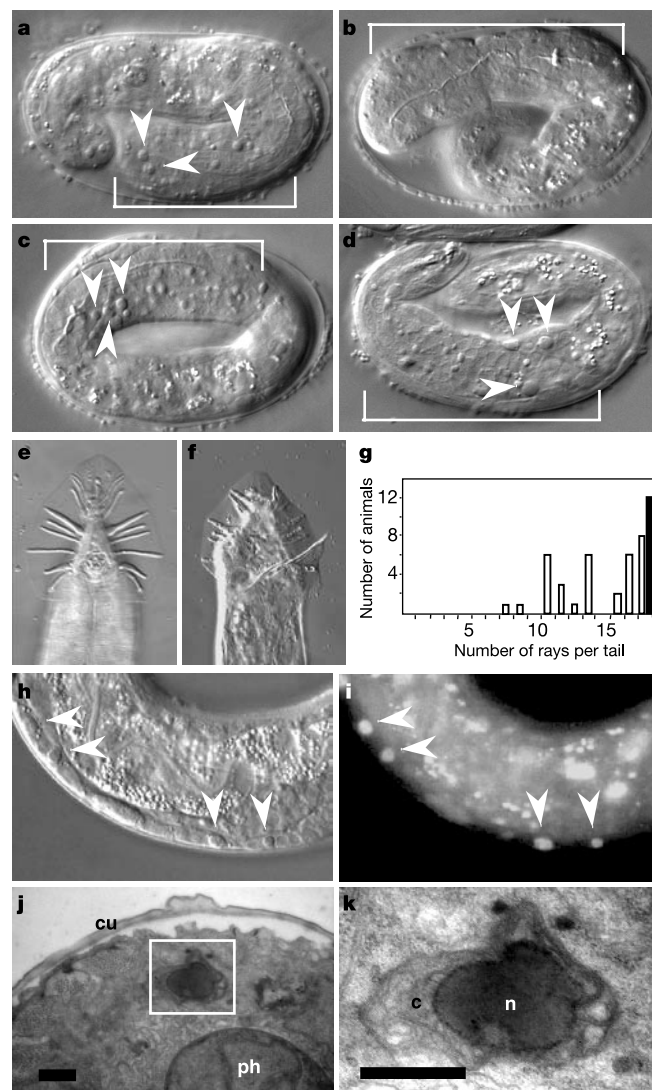


Figure 5 *icd-1(RNAi)*-triggered apoptosis is dependent on CED-4 but not CED-3. **a**, *icd-1(RNAi)* embryo at the threefold stage of development (~400 m.p.f.) with normal and abnormal cell corpses (arrowheads) in the nerve ring region (bracket). **b**, *icd-1(RNAi);ced-4(n1162)* embryo at a similar stage containing neither normal nor abnormal cell corpses. **c**, *icd-1(RNAi);ced-3(n717)* embryo containing several normal and abnormal corpses in the nerve ring region. **d**, *icd-1(RNAi);ced-9(n1950gf)* embryo containing corpses in the nerve ring region. **e**, Tail fan of an *icd-1(RNAi);ced-4(n1162)* male containing 18 normal rays. **f**, Tail fan of *icd-1(RNAi);ced-3(n717)* male missing five rays. Of the remaining rays, most are severely deformed. **g**, Quantification of ray number in *icd-1(RNAi)* embryos carrying a *ced-3* (open bar) or *ced-4* (filled bars) mutation. **h–k**, Microscopic images of *icd-1(RNAi);ced-3(n717)* larvae. **h**, Cell corpses (arrowheads) in the ventral nerve cord. **i**, Cell corpses in **h** stained with SYTO 12. **j**, Electron microscopy image of a cell corpse in the nerve ring region. This corpse (box) is located between the cuticle (cu) and pharynx (ph), where neurons of the nerve ring are typically found. **k**, Magnification of the cell corpse shown in **j** identifying regions of the cytoplasm (c) and nucleus (n). The dead cells show characteristic features of canonical apoptosis, including a swollen nucleus containing electron-dense, condensed chromatin and membrane-bound cytoplasmic fragments²⁹. Scale bars, ~0.5 μ m.

examined whether this inappropriate apoptosis involves the three core components of the apoptotic programme, CED-3, CED-4 and CED-9. We found that CED-4, which is essential for apoptosis in *C. elegans*³, is necessary for the increased apoptosis in *icd-1(RNAi)* animals (Fig. 5): a strong *ced-4* loss-of-function mutation eliminates virtually all apoptosis in these mutants, including the corpses that are abnormal in size and shape (Fig. 5b and Table 1).

Three observations indicate that this effect is not due to a failure of *ced-4* mutants to undergo RNAi. First, *ced-4* mutants showed quantitatively the same susceptibility to RNAi of an unrelated gene (*unc-22*, not shown) as did wild-type animals. Second, the embryonic lethality associated with *icd-1(RNAi)* is not substantially different in wild-type and *ced-4* mutants (Table 1). Last, the proportion of early arresting (~20-cell stage) embryos in both populations was unaffected by the *ced-4* mutation (data not shown). This latter finding also underscores the notion that ICD-1 may have another essential role in development that is unrelated to its action in apoptotic suppression. In addition to eliminating the excess cell corpses, the *ced-4* mutation also restores the lost cells in *icd-1(RNAi)* animals. In sharp contrast to *icd-1(RNAi)* males (Fig. 2e, f), all *icd-1(RNAi);ced-4(n1162)* males ($n = 12$) contained the full complement of 18 rays (Fig. 5e, g) and 215 of the 216 individual rays observed appeared normal.

Although CED-4 is essential for nearly all apoptosis in *icd-1(RNAi)* mutants, we found that CED-3 is not. This finding was unexpected because *ced-3* and *ced-4* have been previously thought to have inseparable functions in apoptosis. Early in embryogenesis, most cell deaths observed in *icd-1(RNAi)* animals, including the apparently excess deaths, require CED-3: less than 5% of *icd-1(RNAi);ced-3(n717)* comma-stage embryos ($n = 102$) contained corpses, in contrast to 100% in the corresponding *ced-3(+)* embryos ($n > 100$). Later in development, however, the inappropriate apoptosis becomes largely independent of CED-3. About 50% of *icd-1(RNAi);ced-3(n717)* embryos at the threefold stage (~400 m.p.f.) contained several cell corpses, which is comparable to the roughly 50% observed for *icd-1(RNAi);ced-3(+)* embryos (Fig. 5a, c, and Table 1) and there was no conspicuous decrease (we estimate no more than a 20% difference) in corpse number in these embryos. These corpses are indistinguishable from normal apoptotic corpses on the basis of morphology by Nomarski optics and electron microscopy, and staining with SYTO 12. (Fig. 5c, h–k). A similar result was obtained with a second *ced-3* allele (*n1286*), confirming the CED-3 independence of this apoptosis. The finding that apoptotic corpses are present in a *ced-3* mutant confirms that the increased cell corpse number in *icd-1(RNAi)* animals is a result of inappropriate apoptosis rather than a cell corpse engulfment defect.

This *ced-3*-independent apoptosis is correlated with the loss of several types of cell, particularly neurons, later in development. A third of *icd-1(RNAi);ced-3(n717)* mutant larvae ($n = 51$) showed between two and eight corpses in their ventral nerve cords at any given time, and corpses were also frequently observed in the nerve ring (Fig. 5h–k). In addition, the subnormal number of rays with normal morphology in *icd-1(RNAi);ced-3(n717)* males was

comparable to that in *icd-1(RNAi);ced-3(+)* males (9.7 ± 6.1 versus 10.4 ± 6.0 wild-type rays, respectively; Fig. 5f) and the total number of rays was not appreciably increased by the *ced-3* mutation, in marked contrast to the effect of the *ced-4* mutation (Fig. 5f, g).

A gain-of-function mutation in *ced-9* prevents normal apoptosis, presumably through constitutive repression of CED-4 by CED-9 (ref. 3). We found that the apoptosis observed in *icd-1(RNAi)* is suppressed only weakly by a *ced-9(gf)* mutation (Fig. 5d and Table 1). One explanation for this finding is that some CED-4 may remain free from CED-9 in the *ced-9(gf)* mutant and a much smaller amount of uncomplexed CED-4 may be required to kill cells in the absence of ICD-1 than is required in its presence.

Our results functionally separate *ced-3* from *ced-4*, casting CED-4 as a putative activator of a CED-3-independent apoptotic pathway that is suppressed by ICD-1.

Immunoreactive ICD-1 is first detected as early as the two-cell stage and continues to be expressed widely throughout the lifespan of *C. elegans* (data not shown). In post-embryonic development and adulthood, it is particularly highly expressed in the nerve ring and pre-anal ganglion, as well as in intestinal and germ cells of both hermaphrodites and males. The ventral nerve cord, epidermis and muscles contain smaller amounts of ICD-1, and expression in other cells (such as in the pharynx) is relatively weak. Beginning early in embryogenesis, ICD-1 is present in a web-like and punctuate pattern in the cytoplasm, coincident with the pattern detected by MitoTracker Red, a marker of mitochondria (Fig. 6b–d). ICD-1 is 63% identical and 79% similar to human β NAC (also known as

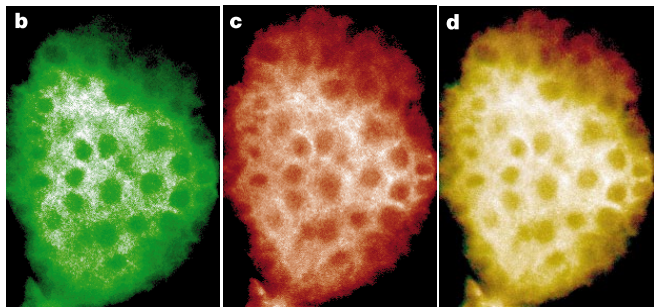
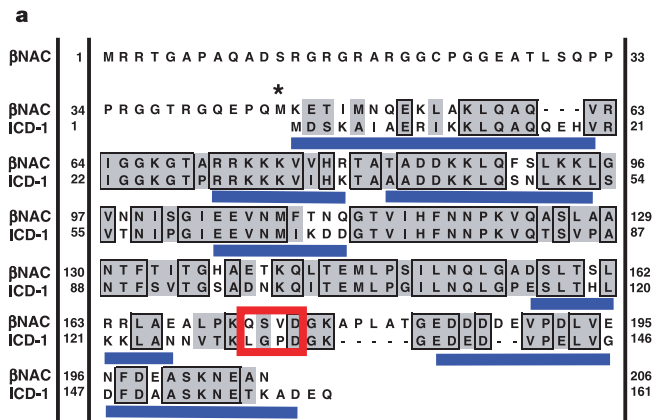


Figure 6 ICD-1 is the homologue of human β NAC and colocalizes with mitochondria. **a**, Amino acid comparison of ICD-1 and human β NAC. There are two isoforms of β NAC, **a** and **b**. The **b** isoform¹⁵ begins at Met 45 (asterisk). Shaded amino acids are similar, and shaded boxed amino acids are identical. The putative caspase cleavage site in ICD-1 (LQPD) and the confirmed caspase cleavage site in β NAC (QSDV)⁶ are shown in the red box, and the six predicted α -helices that constitute a putative CARD¹⁷ in ICD-1 are underlined with blue bars. **b**, Wild-type embryo stained with an antibody specific for ICD-1. **c**, Embryo in **b** stained with MitoTracker Red, a fluorescent marker of mitochondria. **d**, Overlay of **b** and **c**.

Table 1 Effect of mutations in genes involved in the core cell death pathway on apoptosis in *icd-1(RNAi)* embryos

Strain	Type of mutation†	Percentage of embryonic lethality (n)	Embryos with corpses/total embryos* (%)
N2	Wild type	80 (815)	30/65 (46)
<i>ced-4(n1162)</i>	If	72 (234)	1/97 (1)
<i>ced-3(n717)</i>	If	73 (371)	36/77 (47)
<i>ced-9(n1950)</i>	gf	72 (277)	30/102 (29)

**icd-1(RNAi)* embryos were scored at the threefold stage of embryogenesis for the presence of normal and abnormal cell corpses. Embryonic lethality is shown as a measure of RNAi efficiency.

†gf, gain of function; If, loss of function.

BTF3; Fig. 6a), which has been implicated in regulating protein localization during translation¹⁵. This mitochondrial localization of ICD-1 is consistent with the potential role of NAC in stewarding nascent polypeptides to mitochondria during translation¹⁵.

Given the potent anti-apoptotic activity of ICD-1 (Fig. 4), there is likely to be a mechanism that inactivates its function during apoptosis. Indeed, human β NAC is one of the few proteins that is eliminated in Burkitt's lymphoma and Jurkat T cells within 6 h of inducing apoptosis^{5,6}. In addition, β NAC is an *in vitro* substrate of caspase-3 (ref. 6). Structural features of ICD-1 suggest that it interacts with core apoptotic proteins. ICD-1 contains a putative caspase-cleavage site (LGPD; Fig. 6a), similar to other known caspase-cleavage sites (such as LGTD of PKN¹⁶), that is positionally conserved in β NAC at the site cleaved by caspase 3 (ref. 6). In addition, ICD-1 and human β NAC contain a putative caspase recruitment domain (CARD; Fig. 6a)—a set of α -helices that direct interactions with caspases¹⁷—suggesting that it may interact with a caspase or other CARD-containing proteins, such as CED-4/Apaf1 (ref. 18). These structural elements raise the possibility that ICD-1, like human β NAC⁶, may be a caspase substrate.

We have identified a previously uncharacterized suppressor of apoptosis that is active in both differentiated and undifferentiated cells and is highly similar to human β NAC. Past genetic screens for apoptotic components in *C. elegans* have been restricted primarily to mutations without pleiotropic effects (such as lethality) and therefore may have missed apoptotic proteins with other essential cellular functions, such as ICD-1. The intermediate phenotypes that are often observed when using RNAi may have allowed us to distinguish the role of ICD-1 in apoptosis from its other roles in development. RNAi experiments have also allowed us to identify ICD-1 as a specific suppressor of CED-4-mediated, CED-3-independent apoptosis in differentiated cells, raising the possibility of a previously unknown apoptotic pathway in *C. elegans* involving the activation of a non-CED-3 caspases by CED-4. The *C. elegans* genome contains three other apparent caspase-encoding genes that produce six caspase-like proteins¹⁹. Although no *in vivo* function has been assigned to any of these caspases, one of them, CSP-1B, possesses *in vitro* caspase cleavage activity¹⁹. ICD-1 may be required to repress the function of multiple caspases, each of which is capable of activating apoptosis in its absence. □

Methods

Strains, culture conditions and general procedures

N2, *ced-3(n171)*, *ced-3(n1286)*, *ced-4(n1162)*, *ced-9(n1950)* and *ced-1(e1735)* have been described²⁰. The *edl50;him-8* strain was supplied by D. Pilgrim. This strain contains an integrant of a GFP construct driven by the promoter from the F25B3.3 gene, which encodes a putative rasGEF that is expressed throughout the nervous system (D. Pilgrim, personal communication). The *elt-2::GFP* strain was supplied by J. McGhee. We maintained strains by standard procedures²¹. Molecular manipulations were done by standard methods²².

RNAi experiments

We carried out RNAi experiments by using described feeding, soaking and injection protocols^{23–25}. For the feeding protocol, animals were grown at 15°C. For soaking and injections, *icd-1* double-stranded RNA was generated with Megascript T7 (Ambion) from a template amplified by polymerase chain reaction. For feeding, *icd-1* double-stranded RNA was generated in bacteria from an isopropyl-1-thio- β -D-galactopyranoside (IPTG)-inducible plasmid constructed as described²³.

Heat-shock experiments

We expressed *icd-1* under the control of the *hsp-16-2* and *hsp-16-41* promoters by cloning the *icd-1* complementary DNA into vectors pPD49.78 and pPD49.83 (supplied by A. Fire). The resulting plasmids were microinjected into N2 animals and transgenic arrays were integrated by using ultraviolet irradiation. Gravid adults from the resulting strain (JR2028) were dissected and the isolated embryos were given a heat shock at 33°C for 60 min. We scored comma-stage embryos for the presence of cell corpses after 3–4 h. Heat-shocked embryos were allowed to develop to the L3 or L4 larval stage, and extra cells were scored in the anterior pharynx⁸. The heat-shock effect was observed with two independent transgenic lines.

Electron microscopy

We fixed *icd-1(RNAi);ced-3(n171)* animals and embedded for electron microscopy by

standard procedures²⁶, except that no aldehyde fixation or *en block* staining was performed.

Nuclear and cell corpse counts

The number of cell corpses in embryos and larvae counted at the comma stage of embryogenesis, extra cells counted in the anterior pharynx of larvae, and rays in males tails were determined by using Nomarski optics as described^{8,12}. Cell corpses were scored with SYTO 12 staining by placing larvae in 33 μ M SYTO 12 (Molecular Probes) for 3–4 h, allowing them to clear their guts of bacteria on an unseeded plate for 1 h, and visualizing staining by fluorescence microscopy. For corpse quantification, only corpses of a normal size and shape were counted. Because both normal corpses and corpses of abnormal shape and size were eliminated in a *ced-4* mutant, these counts are underestimates of the amount of cell death occurring as a result of attenuated *icd-1* activity. Total cells in the pharynx anterior to and including the retro-vesicular ganglion were counted after staining by DAPI (Vector Labs) using standard procedures²⁷. All data are reported as the mean $\pm$ s.d.

Antibody preparation and immunostaining

After identification through sequence searching software (DNASTAR), two ICD-1-based peptides were synthesized by United Biochemical Research and used to generate two separate polyclonal antibodies (Cocalico). Affinity purification was done with a Sulfolink kit (Pierce). For antibody staining, embryos, larvae or adults were fixed by standard methods²⁷. Samples were disrupted by the freeze–crack method, fixed in methanol and acetone, and blocked with 5% bovine serum albumin in PBS. Slides were incubated with antibodies against ICD-1 for 2 h at room temperature (22°C), washed with PBS plus 0.5% Tween (Sigma) (PBST), incubated with fluorescein isothiocyanate (FITC)-conjugated goat secondary antibodies against rabbit (Sigma) for 1 h at room temperature, and then washed in PBST. We mounted samples with Vectashield mounting medium (Vector Labs) and observed the fluorescence signal. For mitochondrial colocalization experiments, 10 nM MitoTracker Red 580 (Molecular Probes) was added for 5 min after staining with antibodies against ICD-1, and samples were washed in PBST before mounting with Vectashield.

Received 6 January; accepted 22 July 2003; doi:10.1038/nature01920.

- Yuan, J. & Yankner, B. A. Apoptosis in the nervous system. *Nature* **407**, 802–809 (2000).
- Williams, G. T. Programmed cell death: apoptosis and oncogenesis. *Cell* **65**, 1097–1098 (1991).
- Liu, Q. A. & Hengartner, M. O. The molecular mechanism of programmed cell death in *C. elegans*. *Ann. NY Acad. Sci.* **887**, 92–104 (1999).
- Ravagnan, L., Roumier, T. & Kroemer, G. Mitochondria, the killer organelles and their weapons. *J. Cell Physiol.* **192**, 131–137 (2002).
- Brockstedt, E., Otto, A., Rickers, A., Bommert, K. & Wittmann-Liebold, B. Preparative high-resolution two-dimensional electrophoresis enables the identification of RNA polymerase B transcription factor 3 as an apoptosis-associated protein in the human BL60-2 Burkitt lymphoma cell line. *J. Protein Chem.* **18**, 225–231 (1999).
- Thiede, B., Dimmler, C., Siejak, F. & Rudel, T. Predominant identification of RNA-binding proteins in Fas-induced apoptosis by proteome analysis. *J. Biol. Chem.* **276**, 26044–26050 (2001).
- Ellis, H. M. & Horvitz, H. R. Genetic control of programmed cell death in the nematode *C. elegans*. *Cell* **44**, 817–829 (1986).
- Ledwich, D., Wu, Y. C., Driscoll, M. & Xue, D. Analysis of programmed cell death in the nematode *Caenorhabditis elegans*. *Methods Enzymol.* **322**, 76–88 (2000).
- Xu, K., Tavernarakis, N. & Driscoll, M. Necrotic cell death in *C. elegans* requires the function of calreticulin and regulators of Ca^{2+} release from the endoplasmic reticulum. *Neuron* **31**, 957–971 (2001).
- Hersh, B. M., Hartwig, E. & Horvitz, H. R. The *Caenorhabditis elegans* mucopolin-like gene *cup-5* is essential for viability and regulates lysosomes in multiple cell types. *Proc. Natl Acad. Sci. USA* **99**, 4355–4360 (2002).
- Sulston, J. E., Albertson, D. G. & Thomson, J. N. The *Caenorhabditis elegans* male: postembryonic development of nongonadal structures. *Dev. Biol.* **78**, 542–576 (1980).
- Hengartner, M. O., Ellis, R. E. & Horvitz, H. R. *Caenorhabditis elegans* gene *ced-9* protects cells from programmed cell death. *Nature* **356**, 494–499 (1992).
- Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
- Sugimoto, A., Friesen, P. D. & Rothman, J. H. Baculovirus p35 prevents developmentally programmed cell death and rescues a *ced-9* mutant in the nematode *Caenorhabditis elegans*. *EMBO J.* **13**, 2023–2028 (1994).
- Rospert, S., Dubaquitte, Y. & Gautschi, M. Nascent-polypeptide-associated complex. *Cell Mol. Life Sci.* **59**, 1632–1639 (2002).
- Takahashi, M., Mukai, H., Toshimori, M., Miyamoto, M. & Ono, Y. Proteolytic activation of PKN by caspase-3 or related protease during apoptosis. *Proc. Natl Acad. Sci. USA* **95**, 11566–11571 (1998).
- Vaughn, D. E., Rodriguez, J., Lazebnik, Y. & Joshua-Tor, L. Crystal structure of Apaf-1 caspase recruitment domain: an α -helical Greek key fold for apoptotic signaling. *J. Mol. Biol.* **293**, 439–447 (1999).
- Seshagiri, S., Chang, W. T. & Miller, L. K. Mutational analysis of *Caenorhabditis elegans* CED-4. *FEBS Lett.* **428**, 71–74 (1998).
- Shaham, S. Identification of multiple *Caenorhabditis elegans* caspases and their potential roles in proteolytic cascades. *J. Biol. Chem.* **273**, 35109–35117 (1998).
- Hodgkin, J. in *C. elegans* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 881–1047 (Cold Spring Harbor Laboratory Press, Plainview, NY, 1997).
- Lewis, J. A. & Fleming, J. T. in *Caenorhabditis elegans: Modern Biological Analyses of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 4–27 (Academic, San Diego, 1995).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).
- Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2**, Research0002 (<http://genomebiology.com/2000/2/1/research/0002>) (2001).

24. Maeda, I., Kohara, Y., Yamamoto, M. & Sugimoto, A. Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* **11**, 171–176 (2001).
25. Parrish, S., Fleenor, J., Xu, S., Mello, C. & Fire, A. Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol. Cell* **6**, 1077–1087 (2000).
26. Hall, D. in *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 396–436 (Academic, San Diego, 1995).
27. Miller, D. M. & Shakes, D. C. in *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 365–395 (Academic, San Diego, 1995).
28. Fukushige, T., Hawkins, M. G. & McGhee, J. D. The GATA-factor *elt-2* is essential for formation of the *Caenorhabditis elegans* intestine. *Dev. Biol.* **198**, 286–302 (1998).
29. Loo, D. T. & Rillema, J. R. Measurement of cell death. *Methods Cell Biol.* **57**, 251–264 (1998).

Acknowledgements We thank D. Pilgrim for the *edl20* strain; J. White, K. Strohmaier, K. Linberg and G. Lewis for advice on electron microscopy; B. Derry and T. McCloskey for comments on the manuscript; and members of the Rothman laboratory for discussions. Some nematode strains were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. This work was supported by a Cancer Center of Santa Barbara postdoctoral fellowship to T.B. and by grants from the NIH and the March of Dimes Birth Defects Foundation to J.H.R.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to J.H.R. (rothman@lifesci.ucsb.edu).

Nuclear cataract caused by a lack of DNA degradation in the mouse eye lens

Sogo Nishimoto^{1*}, Kohki Kawane^{1,5}, Rie Watanabe-Fukunaga^{1,6}, Hidehiro Fukuyama^{1*}, Yoshiyuki Ohsawa², Yasuo Uchiyama², Noriyasu Hashida⁵, Nobuyuki Ohguro³, Yasuo Tano³, Takeshi Morimoto⁴, Yutaka Fukuda⁴ & Shigekazu Nagata^{1,5,6}

¹Department of Genetics, ²Department of Cell Biology and Neuroscience,

³Department of Ophthalmology, ⁴Department of Physiology and Biosignaling, Osaka University Medical School, Osaka 565-0871, Japan

⁵Laboratory of Genetics, Integrated Biology Laboratories, Graduate School of Frontier Biosciences, Osaka University, Osaka 565-0871, Japan

⁶Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Osaka 565-0871, Japan

* Present addresses: Production Department, Trans Genic Inc., Kumamoto 861-2202, Japan (S.N.); Laboratory of Molecular Genetics and Immunology, The Rockefeller University, New York, New York 10021, USA (H.F.)

The eye lens is composed of fibre cells, which develop from the epithelial cells on the anterior surface of the lens^{1–3}. Differentiation into a lens fibre cell is accompanied by changes in cell shape, the expression of crystallins⁴ and the degradation of cellular organelles^{5,6}. The loss of organelles is believed to ensure the transparency of the lens, but the molecular mechanism behind this process is not known. Here we show that DLAD ('DNase II-like acid DNase', also called DNase IIβ⁸) is expressed in human and murine lens cells, and that mice deficient in the DLAD gene are incapable of degrading DNA during lens cell differentiation—the undigested DNA accumulates in the fibre cells. The DLAD^{−/−} mice develop cataracts of the nucleus lentis, and their response to light on electroretinograms is severely reduced. These results indicate that DLAD is responsible for the degradation of nuclear DNA during lens cell differentiation, and that if DNA is left undigested in the lens, it causes cataracts of the nucleus lentis, blocking the light path.

To identify the DNase responsible for the degradation of nuclear DNA during lens cell differentiation, we first examined by real-time polymerase chain reaction (PCR) the expression of three DNases (DNase II, DLAD and CAD) in various mouse tissues. DNase II is an

acid DNase found in lysosomes that has an essential role in the degradation of nuclear DNA expelled from erythroid precursor cells⁹, and in the degradation of the DNA of apoptotic cells after macrophages engulf them^{10,11}. DLAD is also an acid DNase that has homology to DNase II^{7,8}. DLAD was reported to be expressed in the mouse liver and human salivary glands and lungs, but the physiological role of DLAD remained unidentified. Caspase-activated DNase (CAD) is a neutral DNase, and is responsible for the cell-autonomous DNA degradation in apoptotic cells¹². An analysis by real-time PCR indicated that the CAD messenger RNA was expressed in various mouse tissues, but not in the lens (data not shown). The DNase II mRNA was also ubiquitously expressed, but its expression level in the lens was low. On the other hand, the DLAD mRNA was highly expressed in the mouse lens, and its expression level in the lens was about 10 times higher than that in the liver (Fig. 1a). All the lens fibre cells are nucleated at the beginning of mouse embryogenesis, and begin to lose their nuclei after embryonic day (E)16¹³. Accordingly, the DLAD mRNA level in the lens was low at E14.5, but much increased at the late stage of mouse embryogenesis (Fig. 1b). The DLAD mRNA was detected in human lens cells from patients with senile cataracts (Fig. 1c). An epithelial cell line established from human lens also expressed a significant level of DLAD mRNA. But little expression of DLAD mRNA was found in the other types of cell lines.

To assess the physiological role of DLAD, we generated DLAD^{−/−} mice by gene targeting. The murine DLAD gene is encoded by 6 exons within 15.0 kilobases (kb) of genomic DNA (Fig. 2a). The region encoded by exons 3 and 4 carries two histidine residues conserved between DNase II and DLAD⁷, which could be the enzymatic active sites. A targeting vector was constructed by replacing exons 3 and 4 with the neo gene, and introduced into mouse embryonic stem (ES) cells. ES clones carrying the mutation were identified by PCR, and mice derived from one representative clone were characterized in detail. Heterozygous animals were

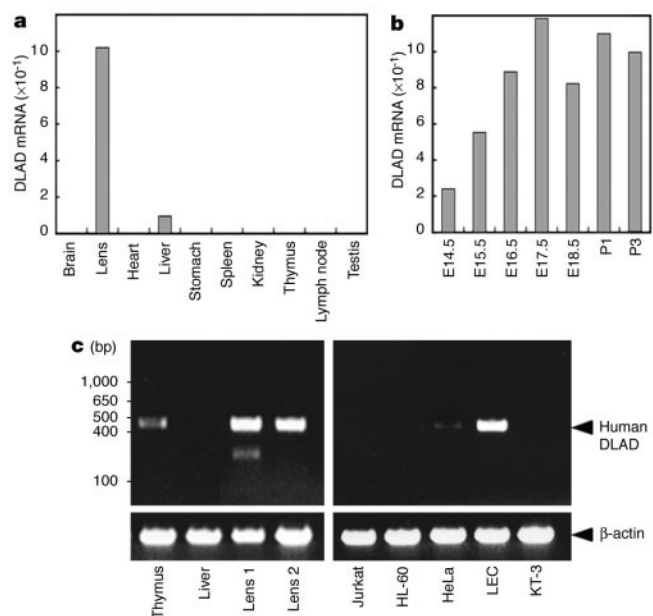


Figure 1 Expression of DLAD mRNA in the lens. **a** and **b**, The level of DLAD mRNA in the indicated mouse tissues (**a**), and in the mouse lens at the indicated developmental stages (**b**), was quantified by real-time PCR, and is expressed as a relative value to β-actin mRNA. **c**, Lens of human cataract patients (Lens 1 and 2), thymus and liver, and cell lines were analysed for DLAD and β-actin mRNAs by RT-PCR. LEC, lens epithelial cells; HeLa, epithelial cells from adenocarcinoma; KT-3 and Jurkat, T cell lymphoma; HL-60, promyelocytic leukaemia.

intercrossed, and the tail DNA of the resulting littermates was analysed by Southern hybridization. As shown in Fig. 2b, *Bam*HI-digested DNA gave a single 9.5-kb band for *DLAD*^{+/+}, a 7.8-kb band for *DLAD*^{-/-}, and both for *DLAD*^{+/-} mice; their sizes were as predicted for the homologous recombination of the *DLAD* gene with the targeting vector. The 1.9-kb *DLAD* mRNA was detected in the lens of *DLAD*^{+/+} mice (Fig. 2c); the size of this mRNA corresponded to that seen in the wild-type liver. The *DLAD*^{-/-} lens also expressed *DLAD* mRNA, but its size was reduced as predicted by the targeting strategy.

To confirm the deletion in the *DLAD* mRNA, RNA from the lens was analysed by PCR with reverse transcription (RT-PCR). As shown in Fig. 2d, a pair of primers that span the full-length coding sequence of *DLAD* mRNA produced a 1,073-base-pair (bp) band with RNA from *DLAD*^{+/+} lens, an 838-bp band for *DLAD*^{-/-} lens, and both for *DLAD*^{+/-} lens. When sequences from exons 4 and 6 were used as primers, RT-PCR with RNA from the *DLAD*^{+/+} and *DLAD*^{+/-} lens produced a band of 636 bp, whereas no bands were detected with RNA from *DLAD*^{-/-} lens. These results confirmed that the *DLAD* mRNA in *DLAD*^{-/-} mice lacked the sequence for exons 3 and 4. The open reading frame from exon 2 was interrupted in exon 5 of the *DLAD*^{-/-} allele, and the truncated *DLAD* protein

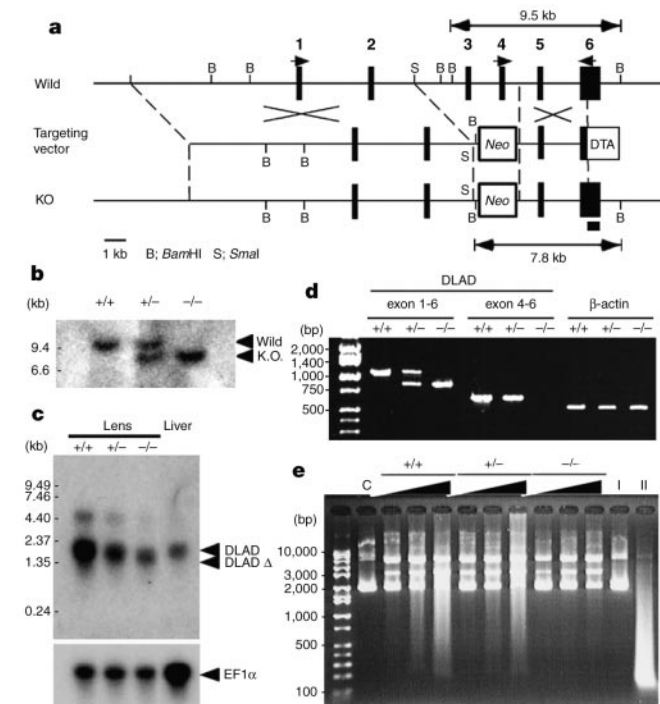


Figure 2 Targeted disruption of the *DLAD* gene. **a**, Endogenous *DLAD* locus (wild), targeting vector, and targeted locus (KO) are shown. Exons are represented by boxes, and are numbered. The neomycin-resistance gene (*Neo*) and the gene coding for the diphtheria toxin A (DTA) are indicated. The probe used for Southern hybridization in **b** is indicated by the box under the lower bar. Primers for RT-PCR in **d** are indicated by arrows above the upper bar. The *Bam*HI-digested fragments detected by the probe are depicted. **b**, Tail DNA (10 μ g) from pups generated from *DLAD*^{+/+} pairs was analysed by Southern hybridization. **c**, Total RNA (10 μ g) from the *DLAD*^{+/+}, *DLAD*^{+/-} and *DLAD*^{-/-} lens, and poly(A) mRNA (2.0 μ g) from the wild-type liver was analysed by northern hybridization with *DLAD* (top) or EF1 α cDNA (bottom). **d**, RNA from *DLAD*^{+/+}, *DLAD*^{+/-} and *DLAD*^{-/-} lenses was analysed by RT-PCR, using primers located on exons 1 and 6, or on exons 4 and 6 of *DLAD* gene. RT-PCR for β -actin mRNA confirms the equivalent amounts of RNA used for RT-PCR. **e**, Plasmid DNA was incubated under acid conditions with the cell extracts (15, 45 and 135 μ g) from *DLAD*^{+/+}, *DLAD*^{+/-} or *DLAD*^{-/-} lens, and analysed by agarose gel electrophoresis. In lane C, DNA was treated without the cell extracts. In lanes I and II, the DNase activities of bovine pancreas DNase I (1.5 ng) and porcine spleen DNase II (3.8 ng) were assayed.

produced in *DLAD*^{-/-} lens was likely to be non-functional. No phenotype in *DLAD*^{+/-} mice (see below) suggests that the truncated protein does not have an adverse effect on the wild-type *DLAD*.

Porcine lens carries an acid DNase called L-DNase II that is derived from a leukocyte elastase inhibitor¹⁴. To examine the existence of acid DNases other than *DLAD* in the mouse lens, and to confirm the null-mutation of the *DLAD* allele in *DLAD*^{-/-} mice, cell extracts were prepared from the lenses, and the DNase activity was assayed. The lens extracts from wild-type and heterozygous mice showed a DNase activity at pH 5.9 (Fig. 2e). On the other hand, little DNase activity was detected under these and more acidic conditions in the lens extracts from *DLAD*^{-/-} mice (Fig. 2e, and data not shown). These results indicated that *DLAD* is the major acid DNase in the mouse lens, and the *DLAD*^{-/-} allele is a null-allele for *DLAD*. The three genotypes (*DLAD*^{+/+}, *DLAD*^{+/-} and *DLAD*^{-/-}) were represented according to mendelian inheritance. Except for the cataract described below, *DLAD*^{-/-} mice showed no gross developmental abnormalities, and were fertile with normal fecundity.

The eyeballs of *DLAD*^{-/-} mice had normal morphology, and were similar in size to the eyes of *DLAD*^{+/+} mice. The lens fibre cells were well differentiated into mature cells, and normal levels of crystallins were found in the lens of *DLAD*^{-/-} mice (data not shown). Histochemical analyses of midsagittal sections of the lens showed a monolayer of cuboidal epithelial cells covering the

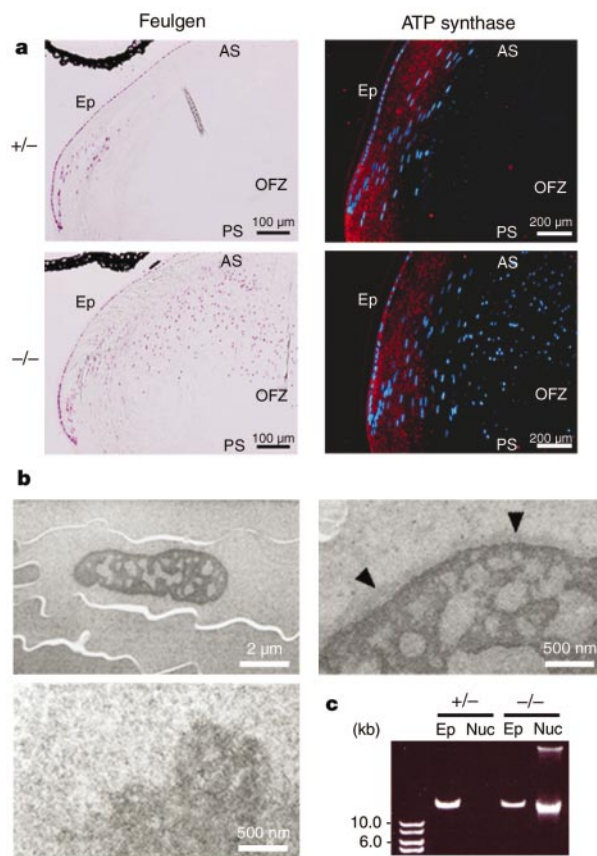


Figure 3 No breakdown of nuclear DNA in *DLAD*^{-/-} lens fibres. **a**, Midsagittal lens slices prepared from *DLAD*^{+/+} and *DLAD*^{-/-} mice were stained with Feulgen, or with the antibody against mitochondrial ATP synthase (red) followed by DAPI staining (blue). Ep, epithelium; OFZ, organelle-free zone; AS, anterior suture; PS, posterior suture. **b**, The lens fibre cells of *DLAD*^{-/-} mice were analysed by electron transmission microscopy. Unpacked materials spreading from the dense bodies are indicated by arrowheads. **c**, DNA from the epithelial layer (Ep) and nucleus (Nuc) of 15-week-old *DLAD*^{+/+} and *DLAD*^{-/-} lens was analysed by electrophoresis on a 0.8% agarose gel.

anterior surface of the lens. Elongating fibre cells at the equator and outer cortex of the lens contained nuclei that stained with Feulgen or DAPI (4,6-diamidino-2-phenylindole; Fig. 3a). They also contained mitochondria, because the cells in this area could be stained by an antibody against mitochondrial ATP synthase. In the inner cortex of the DLAD^{+/+} and DLAD^{+/-} lens, there was an abrupt loss of nuclei, and mitochondria, which formed an 'organelle-free zone'. In contrast, Feulgen- and DAPI-positive materials were found even in the mature fibre cells located in the nucleus of DLAD^{-/-} lens, although these cells seemed to be well differentiated morphologically. These Feulgen-positive materials were found in nuclear lens fibre cells of the young (one week old) or aged (35 week old) DLAD^{-/-} mice (data not shown).

Analyses of the mature lens cells by transmission electron microscopy confirmed that the mature DLAD^{-/-} fibre cells did not contain endoplasmic reticulum or mitochondria (Fig. 3b). But they contained materials that could be condensed DNA. A higher magnification showed that the condensed materials were associated by loosely-packed substances, and some lens fibre cells had amorphous materials that could be unpacked DNA. These materials were not surrounded by membrane-like structures, suggesting that the nuclear membranes were destroyed. To confirm that DNA was left undigested in the mature fibre cells in the DLAD^{-/-} lens, lenses were physically dissected into three regions (epithelial layers, peripheral cortex, and nucleus), and chromosomal DNA was prepared from the cells in epithelial layer and nucleus. As shown in Fig. 3c, the cells in the epithelial layer but not in the nucleus from DLAD^{+/+} lens contained DNA of a large size. In contrast, in the DLAD^{-/-} lens, such DNA was found not only in the cells of the epithelial layer but also in the nucleus.

The eyes of DLAD^{-/-} mice showed a weak but recognizable cataract (Fig. 4a). That is, there was visible opacity in the nucleus of

the DLAD^{-/-} lens, and this was enhanced in aged mice. To examine the effect of the DNA that was left in the fibre cells on the transmission of light, electroretinograms (ERGs) were compared between the DLAD^{+/+} and DLAD^{-/-} mice. As shown in Fig. 4b, when ERGs were recorded from wild-type mice under dark-adapted conditions, typical responses composed of a- and b-waves were observed. In DLAD^{-/-} mice, the responses to the light flash were severely reduced, indicating that the light path was blocked in the DLAD^{-/-} lenses.

In mammals, erythrocytes and lens fibre cells have no nuclei; they are removed during the differentiation into mature cells. We previously reported that the DNA of the nuclei expelled from erythroid precursor cells was cleaved by DNase II in macrophages after the nuclei were engulfed by the macrophages⁹. Here, we have shown that DLAD, a DNase II-like molecule, is responsible for the degradation of DNA in the lens fibre cells. As found for DNase II¹⁵, DLAD is active under acid conditions and seems to be present in lysosomes⁷. However, unlike the DNase II-null mice in which undigested DNA from erythroid precursor cells accumulates in the lysosomes of macrophages^{9,11}, the undigested DNA was found in the fibre cells of the DLAD^{-/-} lens, indicating that DLAD works cell-autonomously to degrade the DNA of fibre cells. This agrees with the fact that the lens is composed only of fibre cells, and no macrophages are present in the lens. Lens fibre nuclei often become TUNEL-positive as they degenerate, suggesting that apoptotic process is involved in this process^{6,16-18}. This apparently contradicts with the biochemical property of DLAD that generates DNA fragments carrying 3'-phosphate⁷. But, there is a report in which DNA fragments carrying free 3'-hydroxyl were not detected in the degenerating fibre cells¹⁹. It is possible that under some circumstances, phosphatase(s) may remove 3'-phosphate from the DLAD-generated DNA fragments, as found for the apoptotic DNA fragments in phagocytes²⁰. Degradation of organelles is often mediated by its own lysosomes through a process called autophagy^{21,22}. However, an involvement of lysosomes in the degradation of nuclei of the lens fibre cells has not been reported^{13,23}, and it is not clear how DLAD works in this process. The human lens epithelial cell line that expresses DLAD may help to elucidate the molecular mechanism of the enucleation process of lens fibre cells.

Finally, the development of cataract in the DLAD^{-/-} mice indicates that DNA must be degraded to ensure the light path in the lens. DLAD is expressed in human lens cells as well. It is possible that some human cataract patients carry a deficiency in the DLAD gene, and the DLAD^{-/-} mice established here will provide a good model system for human cataract. □

Methods

Targeted disruption of the DLAD gene

The DLAD chromosomal gene was isolated from a 129/sv mouse λ gene library and the targeting vector was constructed by replacing the 6.1-kb DNA fragment carrying exons 3–4 with a 2.1-kb neo gene (Fig. 2a). The 1.2-kb DNA fragment coding for the diphtheria toxin A fragment was inserted downstream of the DLAD gene for negative selection. To produce DLAD-deficient mice, ES cells were transfected with the targeting vector as described⁹, and G-418-resistant clones were screened for homologous recombination by PCR. The ES clones carrying the DLAD-deficient allele were used to produce DLAD^{+/+} mutant mice. All mice were housed in a specific pathogen-free facility at Osaka University Medical School, and all animal experiments were carried out in accordance with protocols approved by the Osaka University Medical School Animal Care and Use Committee.

PCR, Southern and northern blot analyses

Genomic DNA was prepared as described²⁴, and the genotype of the DLAD gene was determined by PCR. A sense primer specific for wild-type (5'-AGACAGGGAAGGCAAA TCGACAATAGGGA) or mutant allele (5'-GATTTCGACAGCGCATCGCCTTCTATCG; a sequence in the neomycin-resistant gene) was used with a common antisense primer (5'-CTGCATGCCTTCTGCATTCTGATGCAAG). For Southern hybridization, DNA was digested with *Bam*HI, separated by electrophoresis on a 0.7% agarose gel, transferred to a nitrocellulose filter, and hybridized with a 0.5-kb DNA fragment that is located in exon 6 of the DLAD gene (Fig. 2a). For northern hybridization, RNA was separated on a 1.0% agarose gel, transferred to a nitrocellulose membrane, and hybridized under high stringency conditions as described²⁵.

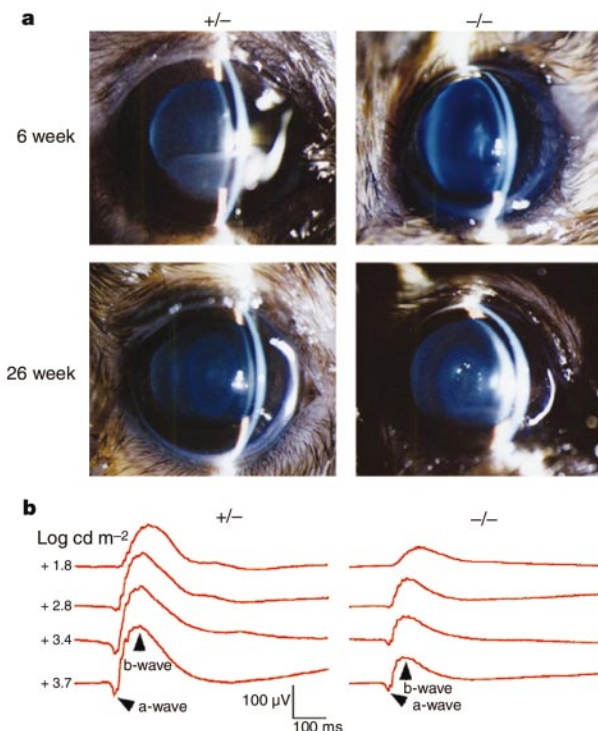


Figure 4 Development of cataract in DLAD^{-/-} mice. **a**, Nuclear cataract in DLAD^{-/-} mice. The eyes of DLAD^{+/+} and DLAD^{-/-} mice at the ages of 6 and 26 weeks were photographed. **b**, Electroretinograms in response to diffuse stimuli of increasing illumination intensities were recorded with 11-week-old DLAD^{+/+} and DLAD^{-/-} mice. The numbers on the left indicate the values of the neutral density (ND) filter in log units.

RT-PCR and real-time PCR

Human lens cells were obtained with informed consent from the anterior capsules of patients with senile cataracts. Human lens epithelial cell line (LEC) was established by infecting lens epithelial cells with Adeno 12-SV40 hybrid virus²⁶, and was previously described²⁷. RNA was prepared by the acid phenol-guanidine thiocyanate method. RNA from human liver and thymus was purchased from Biochain Institute and BD Biosciences Clontech, respectively. The human DLAD mRNA⁸ was detected by RT-PCR using primers 5'-GTGGCATCTGCATAACTTTC and 5'-GGTCTCCAATACATGTCCAG. For real-time PCR, RNA was treated with RNase-free DNase I (Invitrogen), and reverse-transcribed using Superscript II reverse-transcriptase (Invitrogen). The real-time PCR was carried out using a LightCycler (Roche Diagnostics) according to the instructions provided by the manufacturer. Primers for the mouse DLAD mRNA were: 5'-CCAGTTCATGGCTATGAG TAC and 5'-TTAGGTCTCCAATGCAGGTCCAGCGATTTC. As a control, the expression of the mouse β -actin gene was detected using primers 5'-TGTGATGGTGGGAATGGGT CAG and 5'-TTTATGTACACGACGATTTC. The level of specific mRNA was quantified at the point where the LightCycler System detected the upstroke of the exponential phase of PCR accumulation, and normalized to the level of β -actin in each sample.

Assay for acid DNase

The lenses were homogenized in lysis buffer (10 mM Tris-HCl (pH 7.5), 1 M NaCl, and 1 mM PAPMSF). After removing the cell by centrifugation, the supernatant was dialysed against 50 mM citrate buffer (pH 5.0) containing 1 mM EDTA, and the acid DNase activity was measured as described⁹. In brief, plasmid DNA (2.0 μ g) was mixed with the sample in 120 μ l of the reaction mixture (50 mM MOPS-NaOH buffer (pH 5.9) containing 10 mM EDTA). After incubation at 37 °C for 3 h, the mixture was treated with phenol-chloroform, and the DNA was recovered by EtOH precipitation.

Histochemical and electron microscopic analysis of eye lens

For histological analysis, mouse lens at 6 weeks of age was fixed in 4% paraformaldehyde/4% sucrose in 0.1 M phosphate buffer (pH 7.2), embedded in paraffin, and sectioned at 4 μ m. The sections were treated for 10 min with 1 N HCl at 60 °C, and subjected to Feulgen staining. To detect mitochondrial ATP synthase, the sections were incubated overnight at 4 °C with rabbit antibody against subunit β of mitochondrial ATP synthase²⁸, followed by the Cy3-coupled goat anti-rabbit IgG, F(ab')₂ fragment (Jackson ImmunoResearch Lab.). For electron microscopy, the lenses from 6-week-old mice were fixed with 4% glutaraldehyde in 0.1 M phosphate buffer (pH 7.2), post-fixed with 1% OsO₄ at 4 °C for 2 h, and embedded in Epon 812. Sections (80 nm) were prepared with an ultramicrotome (Reichert Ultracut N), stained with lead citrate and uranyl acetate, and observed with an H-7100 electron microscope (Hitachi High-Technologies Co.).

Electroretinograms

ERGs were obtained as described previously²⁹. In brief, mice were dark-adapted for more than 12 h and anaesthetized by a subcutaneous injection of xylazine and ketamine. The cornea was anaesthetized with 0.4% oxybuprocaine, and the pupils were dilated with 0.5% tropicamide and 0.5% phenylephrine-HCl. ERGs were recorded from the corneal surface of one eye using chlorided silver wire loops. The luminance of the unattenuated light stimulus was 3.7 log(cd m⁻²) on the surface of the eye, and different neutral-density filters were used to serially reduce the stimulus intensities. Responses were amplified at a 10,000 gain with a band-pass frequency setting of 0.08–1,000 Hz (Nihon Kohden, VC-11). Single-flash stimulus intensities were increased from 1.8 log(cd m⁻²) to 3.7 (log cd m⁻²), and divided into 4 steps. Ten responses were averaged with an averager (Nihon Kohden, LEG-1000).

Received 28 March; accepted 4 July 2003; doi:10.1038/nature01895.

- McAvoy, J. W., Chamberlain, C. G., de Iongh, R. U., Hales, A. M. & Lovicu, F. J. Lens development. *Eye* **13**, 425–437 (1999).
- Grainger, R. M., Henry, J. J., Saha, M. S. & Servetnick, M. Recent progress on the mechanisms of embryonic lens formation. *Eye* **6**, 117–122 (1992).
- Piatigorsky, J. Lens differentiation in vertebrates. A review of cellular and molecular features. *Differentiation* **19**, 134–153 (1981).
- Wistow, G. J. & Piatigorsky, J. Lens crystallins: The evolution and expression of proteins for a highly specialized tissue. *Annu. Rev. Biochem.* **57**, 479–504 (1988).
- Bassnett, S. Lens organelle degradation. *Exp. Eye Res.* **74**, 1–6 (2002).
- Bassnett, S. & Mataic, D. Chromatin degradation in differentiating fiber cells of the eye lens. *J. Cell Biol.* **137**, 37–49 (1997).
- Shiokawa, D. & Tanuma, S. DLAD, a novel mammalian divalent cation-independent endonuclease with homology to DNase II. *Nucleic Acids Res.* **27**, 4083–4089 (1999).
- Krieser, R. J., MacLea, K. S., Park, J. P. & Eastman, A. The cloning, genomic structure, localization, and expression of human deoxyribonuclease II β . *Gene* **269**, 205–216 (2001).
- Kawane, K. *et al.* Requirement of DNase II for definitive erythropoiesis in the mouse fetal liver. *Science* **292**, 1546–1549 (2001).
- Krieser, R. J. *et al.* Deoxyribonuclease IIa is required during the phagocytic phase of apoptosis and its loss causes lethality. *Cell Death Differ.* **9**, 956–962 (2002).
- Kawane, K. *et al.* Impaired thymic development in mouse embryos deficient in apoptotic DNA degradation. *Nature Immunol.* **4**, 138–144 (2003).
- Nagata, S., Nagase, H., Kawane, K., Mukae, N. & Fukuyama, H. Degradation of chromosomal DNA during apoptosis. *Cell Death Differ.* **10**, 108–116 (2003).
- Kuwabara, T. & Imaizumi, M. Denucleation process of the lens. *Invest. Ophthalmol. Vis. Sci.* **13**, 973–981 (1974).
- Torriglia, A. *et al.* L-DNase II, a molecule that links proteases and endonucleases in apoptosis, derives from the ubiquitous serpin leukocyte elastase inhibitor. *Mol. Cell Biol.* **18**, 3612–3619 (1998).
- Bernardi, G., (ed. Boyer, P. D.) *The Enzymes*, 271–287 (Academic, New York, 1971).

- Wride, M. A. & Sanders, E. J. Nuclear degeneration in the developing lens and its regulation by TNF α . *Exp. Eye Res.* **66**, 371–383 (1998).
- Dahm, R., Gribbon, C., Quinlan, R. A. & Prescott, A. R. Changes in the nucleolar and coiled body compartments precede lamina and chromatin reorganization during fibre cell denudation in the bovine lens. *Eur. J. Cell Biol.* **75**, 237–246 (1998).
- Ishizaki, Y., Jacobson, M. D. & Raff, M. C. A role for caspases in lens fiber differentiation. *J. Cell Biol.* **140**, 153–158 (1998).
- Chaudun, E. *et al.* DNA strand breakage during physiological apoptosis of the embryonic chick lens: Free 3' OH end single strand breaks do not accumulate even in the presence of a cation-independent deoxyribonuclease. *J. Cell. Physiol.* **158**, 354–364 (1994).
- McIlroy, D. *et al.* An auxiliary mode of apoptotic DNA fragmentation provided by phagocytes. *Genes Dev.* **14**, 549–558 (2000).
- Klionsky, D. J. & Emr, S. D. Autophagy as a regulated pathway of cellular degradation. *Science* **290**, 1717–1721 (2000).
- Klionsky, D. J. & Ohsumi, Y. Vacuolar import of proteins and organelles from the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 1–32 (1999).
- Vrensen, G. F., Graw, J. & De Wolf, A. Nuclear breakdown during terminal differentiation of primary lens fibres in mice: A transmission electron microscopic study. *Exp. Eye Res.* **52**, 647–659 (1991).
- Laird, P. W. *et al.* Simplified mammalian DNA isolation procedure. *Nucleic Acids Res.* **19**, 4293 (1991).
- Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001).
- Fleming, T. P., Song, Z. & Andley, U. P. Expression of growth control and differentiation genes in human lens epithelial cells with extended life span. *Invest. Ophthalmol. Vis. Sci.* **39**, 1387–1398 (1998).
- Imoto, Y. *et al.* Effects of RGD peptides on cells derived from the human eye. *Jpn J. Ophthalmol.* (in the press).
- Ezaki, J., Wolfe, L. S. & Kominami, E. Specific delay in the degradation of mitochondrial ATP synthase subunit c in late infantile neuronal ceroid lipofuscinosis is derived from cellular proteolytic dysfunction rather than structural alteration of subunit c. *J. Neurochem.* **67**, 1677–1687 (1996).
- Masu, M. *et al.* Specific deficit of the ON response in visual transmission by targeted disruption of the mGluR6 gene. *Cell* **80**, 757–765 (1995).

Acknowledgements We thank E. Kominami for providing us with rabbit antibody against the subunit β of mitochondrial ATP synthase. We thank K. Miwa for cloning of mouse DLAD gene, Y. Seto for maintaining the mice, and M. Fujii and M. Harayama for secretarial assistance. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, Sports, and Culture in Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.N. (nagata@genetic.med.osaka-u.ac.jp).

Determining the position of the cell division plane

Julie C. Canman¹*, Lisa A. Cameron¹†, Paul S. Maddox¹†, Aaron Straight², Jennifer S. Tirnauer², Timothy J. Mitchison², Guowei Fang³, Tarun M. Kapoor⁴ & E. D. Salmon¹

¹University of North Carolina, Department of Biology, 607 Fordham Hall, CB #3280, Chapel Hill, North Carolina 27699-3280, USA

²Harvard Medical School, Department of Cell Biology, 250 Longwood Avenue, Boston, Massachusetts 02115, USA

³Stanford University, Department of Biological Sciences, Stanford, California 94305, USA

⁴Laboratory of Chemistry and Cell Biology, The Rockefeller University, New York, New York 10021, USA

* Present address: University of Oregon, Institute of Molecular Biology, 1370 Franklin Blvd., Eugene, Oregon, USA

† These authors contributed equally to this work

Proper positioning of the cell division plane during mitosis is essential for determining the size and position of the two daughter cells—a critical step during development and cell differentiation¹. A bipolar microtubule array has been proposed to be a minimum requirement for furrow positioning in mammalian cells, with furrows forming at the site of microtubule plus-end overlap between the spindle poles^{2–4}. Observations in other species have suggested, however, that this may not be true^{5,6}. Here we show, by inducing mammalian tissue cells with

monopolar spindles to enter anaphase^{7,8}, that furrow formation in cultured mammalian cells does not require a bipolar spindle. Unexpectedly, cytokinesis occurs at high frequency in monopolar cells. Division always occurs at a cortical position distal to the chromosomes. Analysis of microtubules during cytokinesis in cells with monopolar and bipolar spindles shows that a subpopulation of stable microtubules extends past chromosomes and binds to the cell cortex at the site of furrow formation. Our data are consistent with a model in which chromosomes supply microtubules with factors that promote microtubule stability and furrowing.

How microtubules control the positioning of the cell division plane is the subject of much controversy. In current models, microtubules either promote furrowing at the cell equator by means of the overlap zone of microtubules from opposite poles, or restrict furrowing to the cell equator by means of the microtubule asters inhibiting cortical contractility at the cell poles²⁻⁴. A central tenet of these models is that dividing cells must have bipolar arrays of microtubules.

We used monastrol, an inhibitor of the kinesin Eg5 and centrosome separation^{7,8}, to test whether a bipolar microtubule array is essential for cytokinesis in mammalian tissue cells. Treatment of PtK1 cells with monastrol resulted in both the formation of monopolar half-spindles with two unseparated centrosomes near one spindle pole (Fig. 1a) and activation of the mitotic spindle checkpoint. To overcome the spindle checkpoint and induce the cytokinetic phase (C-phase)⁹ of the cell cycle, cells with monopolar spindles were injected with anti-Mad2 antibodies, Mad2ΔC, or Mad1F10 (refs 10–12). After injection, cells with monopolar spindles entered anaphase in 22 ± 7 min (Fig. 1b, c, and Supplementary Movie 1) and both sister chromatids moved toward the spindle pole ($n = 15$, Fig. 1c and Supplementary Movie 1).

More than 90% of the cells that entered anaphase underwent cytokinesis (14/15, Fig. 1c and Supplementary Movie 1). Furrows formed within 6 ± 2 min, a time that is typical of control cells with bipolar spindles⁹. Monopolar spindles were frequently asymmetric and resembled a 'half-spindle' with all chromosomes on one side of the unseparated centrosomes (Fig. 1a). Likewise, cytokinesis was polarized in these cells: cytokinetic furrows always occurred on the

side of the cell facing the chromosomes and not on the side facing the unseparated centrosomes (14/14; Fig. 1c and see below). Furrowing occurred $25 \pm 4 \mu\text{m}$ away from the spindle pole, resulting in the formation of a small anucleate cytoblast (Fig. 1c). These cytokinetic furrows were stable, persisting for at least 3 h after furrow completion and cell respreading. Thus, two opposing microtubule arrays are not essential for cytokinesis in cultured mammalian cells.

In normal cells 'chromosome passenger' proteins, including INCENP, Survivin and Aurora B, translocate along microtubules during anaphase and cytokinesis from the centromeric chromatin to the cell equator¹²⁻¹⁵. An accumulation of passenger proteins at the equator is essential for cytokinesis¹⁶ and is thought to depend on an antiparallel microtubule array¹⁵. To test whether centromere-associated proteins relocate to the furrow region of cells with monopolar spindles, we examined changes in the localization of INCENP after anaphase onset¹². Monastrol-treated PtK1 cells with monopolar spindles were fixed either before or after microinjection with Mad2ΔC at the time of furrow initiation. Cells were then immunostained for tubulin, INCENP and DNA. In non-injected cells with monopolar spindles, INCENP localized to the centromeres (data not shown). In monopolar cells induced into anaphase, there was marked accumulation of INCENP at the furrow (Fig. 2). Thus, an overlapping bipolar array of microtubules is not needed to localize INCENP to the furrow region.

As microtubules are essential for furrow positioning⁹, we wanted to examine the behaviour of microtubules during cytokinesis in living cells with monopolar spindles. We pretreated cells with monastrol, co-injected fluorescently labelled tubulin and Mad2ΔC, and imaged microtubule dynamics by spinning disk confocal microscopy. At anaphase onset, the microtubules lengthened markedly, growing out towards the cell cortex (Fig. 3a). Although many of these microtubules were highly dynamic, a subset of microtubules were stable and bound to the cell cortex. These stable microtubules were found only on the side of the cell facing the chromosomes. Kymographic analysis indicated that cortical microtubules extending past the chromosomes were four times as stable as microtubules growing towards the side of the cell facing the spindle pole (median lifetime 224 ± 119 s ($n = 115$) at furrow versus

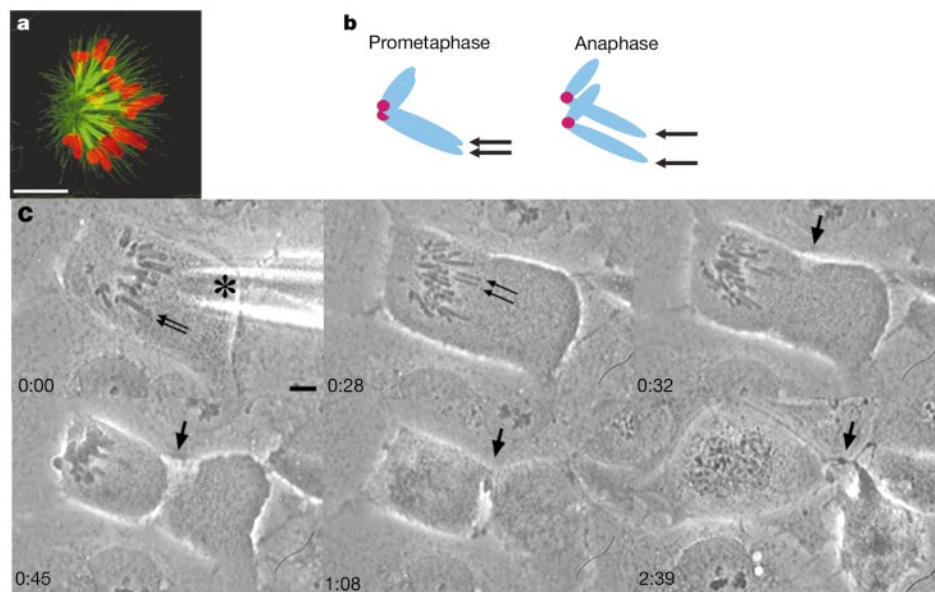


Figure 1 Cytokinesis in cells with monopolar spindles **a**, Monastrol-treated PtK1 cell in prometaphase cell stained for tubulin (green) and DNA (red). **b**, Diagram of anaphase in monastrol with both sister chromatids facing the same spindle pole. **c**, Cytokinesis in a

cell induced into anaphase by microinjection of Mad2ΔC. Asterisk indicates injection needle. Small arrows indicate sister separation. Large arrow indicates furrow position. In each frame, time is given in h:min. Scale bars, $5 \mu\text{m}$.

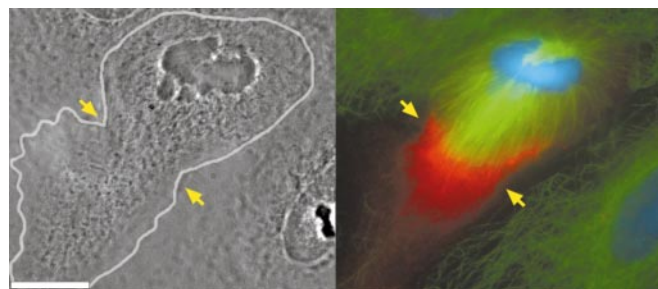


Figure 2 INCENP localizes to furrows in cells with monopolar spindles. Monastrol-treated PtK1 cell injected with Mad2 Δ C, recorded through anaphase and furrow onset, and then fixed. Left, phase contrast highlights cell outline (traced in white). Arrows indicate furrow site. Right, the cell was then processed for immunofluorescence of tubulin (green), INCENP (red) and DNA (blue). Scale bar, 10 μ m.

64 $\pm$ 38 s (n = 187) at pole; Fig. 3b, c, and Supplementary Movie 2). This precise correlation between the position of the chromosomes and the appearance of stable microtubules indicates that chromosomes may have a direct role in furrow positioning through microtubules.

To test further this relationship between chromosomes and stable microtubules, we examined the behaviour of microtubules in cells with bipolar spindles injected with labelled tubulin. After anaphase onset, but before furrow onset, we observed a marked polymerization of microtubules toward the cell cortex in the furrow regions as well as in the polar regions (Fig. 3d and Supplementary Movie 3). Kymographic analysis of these movies showed that dynamic micro-

tubules, which populated the polar regions of the cell, did not form stable cortical interactions even if they contacted the cell cortex at the site of furrow formation (median lifetime 60 $\pm$ 41 s (n = 247); for example, see left aster, bottom equator in Supplementary Movie 3 and Fig. 3e). These microtubules may be acting to suppress furrowing activity in the polar regions of the cell as ectopic furrowing occurs throughout the cell cortex in the absence of microtubules^{9,17,18}.

In contrast, a stable subpopulation of microtubules was located in the region of the furrow (median lifetime, 123 $\pm$ 60 s (n = 207); Fig. 3f and Supplementary Movie 3). Notably, these stable microtubules were associated with, or 'passed by', a kinetochore fibre or chromosome for about 75% of the time (87/115 stable microtubules; Fig. 3g, h). Dynamic microtubules could be also seen in the equatorial region; however, these microtubules did not bind the cell cortex. Thus, in cells with bipolar spindles there is also a precise correlation between the localization of the chromosomes and the formation of stable microtubules in both PtK1 cells (Fig. 3f) and in U2OS cells (data not shown). Therefore, cytokinesis in cells with bipolar spindles (Fig. 4b) seems to occur as the sum of two asymmetric monopolar spindles (Fig. 4a): the stable, chromosome-associated microtubules bind the cell cortex at the site of furrow formation, whereas dynamic astral microtubules populate the polar regions.

We have shown that microtubule dynamics are asymmetric within a single cell during cytokinesis, with furrows forming near areas of high numbers of stable microtubule ends. This is seemingly in opposition to studies in *Caenorhabditis elegans*, which concluded that furrow formation is dependent on a reduced microtubule density in the furrow region¹⁹. It is also possible, however, that a

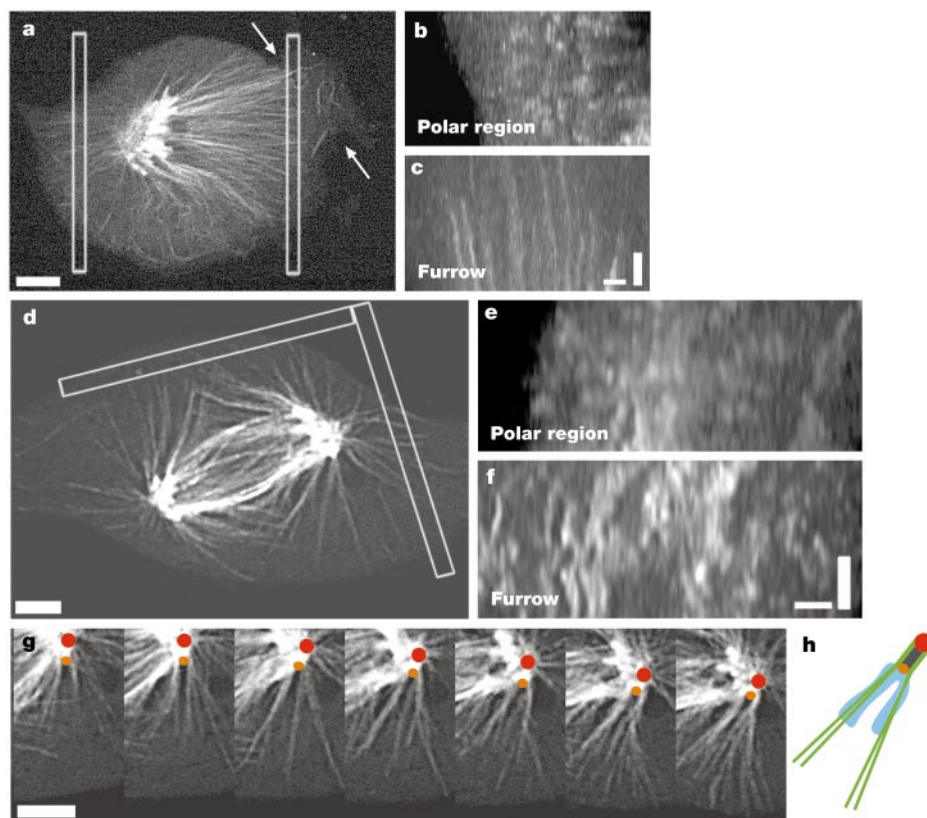


Figure 3 Chromosome-associated stable microtubules at the furrow. **a, d**, Live imaging of labelled tubulin in monopolar spindles (**a**) and untreated bipolar spindles (**d**) during anaphase. White rectangles roughly indicate source for kymographs. **b, c, e, f**, Kymograph of microtubules in the polar region (**b, e**) and furrow (**c, f**) used to

analyse microtubule stability. **g**, Close-up view showing stable microtubules associated with the kinetochore fibres at their minus ends. Red circles mark the end of the kinetochore fibre. Orange circles mark the kinetochore. **h**, Schematic drawing of **g**. Horizontal scale bars, 5 μ m; vertical scale bars, 400 s.

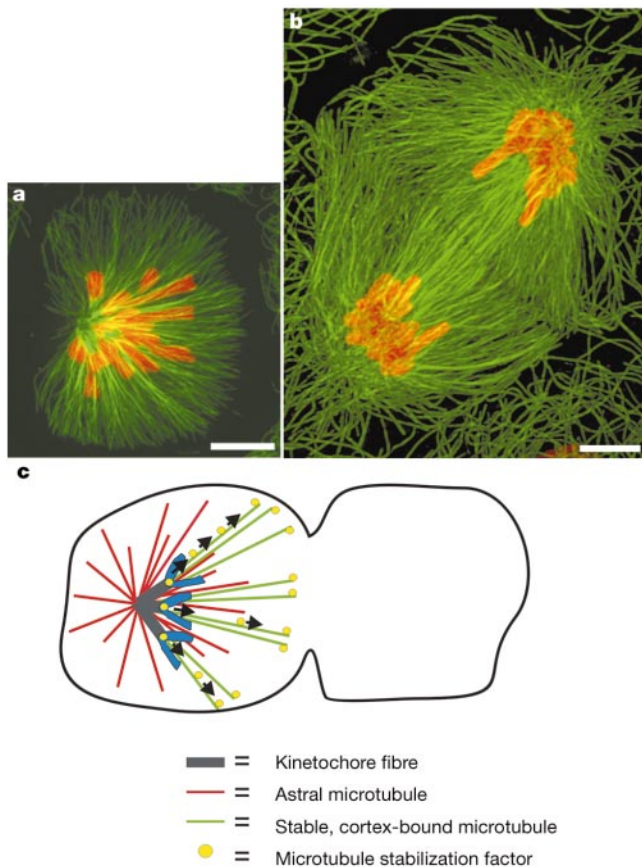


Figure 4 Model of cytokinesis induced by monopolar spindles. **a**, **b**, Normal bipolar microtubule array (**b**) is the sum of two, oppositely oriented, cytokinesis-inducing monopolar spindles (**a**). **c**, Centromere-bound microtubule stabilization factors are transported to the flanking microtubules on anaphase onset.

steep gradient of high numbers of stable microtubules flanking a region of low microtubule density is essential for furrow formation⁴. Models of how microtubules produce a non-uniform distribution of the cleavage stimulus during cytokinesis have previously relied on a difference in the numbers of microtubules in the polar versus the equatorial regions of the cell (for example, see refs 1–3, 4, 19, 20). Our observations suggest instead that a variation in microtubule dynamics in different regions of the cell is responsible for positioning the cell division plane. Specifically, chromosome-associated stable microtubules stimulate contractility in the cell equator. In addition, our data concur with the possibility that non-chromosome-associated dynamic microtubules may prevent ectopic furrowing in the polar regions of the cell, as would occur in the absence of microtubules^{9,17,18}.

In motile cells, differential microtubule dynamics seem to control the actin cytoskeleton differentially. That is, dynamic microtubules at the leading edge of the cell promote actin-dependent cell protrusion and stable microtubules at the cell rear promote actin-dependent cell tail retraction²¹. We propose that a similar asymmetry in microtubule dynamics leads to non-uniformity of a cleavage stimulus resulting in cytokinesis. We have constructed a model of cytokinesis in cultured cells in which microtubules associated with chromosomes or centromeres form persistent associations with the cell cortex through the transport of centromere-bound microtubule stabilization factors, of yet unknown identity, to the plus ends of these microtubules (Fig. 4c). These stable microtubules trigger the cell cortex to become contractile, forming a productive furrow in their vicinity. At the same time, the

dynamic microtubules in the polar regions are probably acting to suppress ectopic furrowing outside the cell equator. This model would explain the furrowing pattern seen in *Sciara coprophila* monopolar meiosis²² and cellularizing syncytial *Drosophila melanogaster* embryos²³, and may be important for generating cells without nuclei, such as platelets²⁴. □

Methods

Tissue culture, drug treatment, and anaphase induction

Monastrol was used at a concentration of 100 μ M in either F-12 medium (Sigma) for pretreatment or phenol-red-free L-15 medium (Sigma) for filming. Pretreatments were done for at least 1 h, but up to 6 h, before imaging. We cultured, injected and imaged PtK1 cells as described⁹. His-tagged Mad2 Δ C and glutathione S-transferase (GST)-tagged Mad1F10 were used to induce precocious anaphase as described^{11,12}.

Injection and filming chambers

Mitotic PtK1 cells were injected with tubulin labelled with X-rhodamine either alone or together with X-phalloidin labelled with Alexa 488 at room temperature to maintain the cells in mitosis throughout the injections. All filming was done at 37 $^{\circ}$ C as described²⁵. For imaging tubulin, we used a needle concentration of 1 μ M tubulin. Chambers were prepared²⁵ using phenol-red-free L-15 (Gibco-BRL) supplemented with Oxyrase (Oxyrase Inc.) to minimize photobleaching.

Live-cell imaging

Stage temperature was maintained at about 35–37 $^{\circ}$ C using an ASI 400 air curtain incubator (Nevtek). Images were taken at intervals of 15–20 s with an UltraVIEW spinning disc confocal scanning head (Perkin-Elmer) on a TE300 microscope (Nikon) using an Orca I, Orca II or Orca ER camera (Hamamatsu Photonics). These images were then digitally processed to enhance the contrast, using Metamorph imaging software (Universal Imaging).

Kinetochore fibre and chromosome identification

Kinetochore fibres were identified as the small bright bundles of microtubules terminating at a chromosome with the kinetochore at the site of the chromosome/kinetochore fibre interface. Chromosomes were easily identifiable because they excluded the free labelled tubulin that filled the remainder of the cell.

Cortex interaction measurements

Custom alignment and rotation algorithms in MetaMorph were used to correct for spindle and/or cell rotation. Aligned sequences were analysed by kymographs. In brief, a region of interest within the sequence (that is, the equatorial region of the cortex) was cropped and entered into the three-dimensional reconstruction algorithm in MetaMorph and a single '90 $^{\circ}$ ' view through the stack was generated. The resulting image was a two-dimensional representation of the time-lapse movie, in which time was the y dimension and positions over time of all the microtubules in this region appeared as lines parallel with the y axis. We obtained lifetimes of microtubules in the selected region by measuring the length of the lines in the kymograph images. Statistical analysis was done in Excel (Microsoft).

Received 31 January; accepted 9 June 2003; doi:10.1038/nature01860.

Published online 6 August 2003.

1. Rappaport, R. *Cytokinesis in Animal Cells* (Cambridge Univ, Cambridge, UK, 1997).
2. Harris, A. K. & Gewalt, S. L. Simulation testing of mechanisms for inducing the formation of the contractile ring in cytokinesis. *J. Cell Biol.* **109**, 2215–2223 (1989).
3. White, J. G. & Borisy, G. G. On the mechanism of cytokinesis in animal cells. *J. Theor. Biol.* **101**, 289–316 (1983).
4. Mandato, C. A., Benink, H. A. & Bement, W. M. Microtubule-actomyosin interactions in cortical flow and cytokinesis. *Cell Motil. Cytoskeleton* **45**, 87–92 (2000).
5. Wolf, N., Hirsh, D. & McIntosh, J. R. Spermatogenesis in males of the free-living nematode, *Caenorhabditis elegans*. *J. Ultrastruct. Res.* **63**, 155–169 (1978).
6. Rappaport, R. Experiments concerning the cleavage stimulus in sand dollar eggs. *J. Exp. Zool.* **148**, 81–89 (1961).
7. Mayer, T. U. et al. Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971–974 (1999).
8. Kapoor, T. M., Mayer, T. U., Coughlin, M. L. & Mitchison, T. J. Probing spindle assembly mechanisms with monastrol, a small molecule inhibitor of the mitotic kinesin, Eg5. *J. Cell Biol.* **150**, 975–988 (2000).
9. Canman, J. C., Hoffman, D. B. & Salmon, E. D. The role of pre- and post-anaphase microtubules in the cytokinesis phase of the cell cycle. *Curr. Biol.* **10**, 611–614 (2000).
10. Gorbisky, G. J., Chen, R. H. & Murray, A. W. Microinjection of antibody to Mad2 protein into mammalian cells in mitosis induces premature anaphase. *J. Cell Biol.* **141**, 1193–1205 (1998).
11. Canman, J. C., Salmon, E. D. & Fang, G. Inducing precocious anaphase in cultured mammalian cells. *Cell Motil. Cytoskeleton* **52**, 61–65 (2002).
12. Canman, J. C. et al. Anaphase onset does not require the microtubule-dependent depletion of kinetochore and centromere-binding proteins. *J. Cell Sci.* **115**, 3787–3795 (2002).
13. Wheatley, S. P., Kandels-Lewis, S. E., Adams, R. R., Ainsztein, A. M. & Earnshaw, W. C. INCENP binds directly to tubulin and requires dynamic microtubules to target to the cleavage furrow. *Exp. Cell Res.* **262**, 122–127 (2001).
14. Murata-Hori, M. & Wang, Y. L. Both midzone and astral microtubules are involved in the delivery of cytokinesis signals: insights from the mobility of aurora B. *J. Cell Biol.* **159**, 45–53 (2002).
15. Adams, R. R., Carmona, M. & Earnshaw, W. C. Chromosomal passengers and the (aurora) ABCs of mitosis. *Trends Cell Biol.* **11**, 49–54 (2001).
16. Mackay, A. M., Ainsztein, A. M., Eckley, D. M. & Earnshaw, W. C. A dominant mutant of inner

- centromere protein (INCENP), a chromosomal protein, disrupts prometaphase congression and cytokinesis. *J. Cell Biol.* **140**, 991–1002 (1998).
17. Shannon, K. B., Canman, J. C. & Salmon, E. D. Mad2 and BubR1 function in a single checkpoint pathway that responds to a loss of tension. *Mol. Biol. Cell* **13**, 3706–3719 (2002).
 18. Kurz, T. *et al.* Cytoskeletal regulation by the Nedd8 ubiquitin-like protein modification pathway. *Science* **295**, 1294–1298 (2002).
 19. Dechant, R. & Glotzer, M. Centrosome separation and central spindle assembly act in redundant pathways that regulate microtubule density and trigger cleavage furrow formation. *Dev. Cell* **4**, 333–344 (2003).
 20. Asnes, C. F. & Schroeder, T. E. Cell cleavage. Ultrastructural evidence against equatorial stimulation by aster microtubules. *Exp. Cell Res.* **122**, 327–338 (1979).
 21. Wittmann, T. & Waterman-Storer, C. M. Cell motility: can Rho GTPases and microtubules point the way? *J. Cell Sci.* **114**, 3795–3803 (2001).
 22. Kubai, D. F. Meiosis in *Sciara coprophila*: structure of the spindle and chromosome behavior during the first meiotic division. *J. Cell Biol.* **93**, 655–669 (1982).
 23. Foe, V. E., Field, C. M. & Odell, G. M. Microtubules and mitotic cycle phase modulate spatiotemporal distributions of F-actin and myosin II in *Drosophila* syncytial blastoderm embryos. *Development* **127**, 1767–1787 (2000).
 24. Radley, J. M. & Scurfield, G. The mechanism of platelet release. *Blood* **56**, 996–999 (1980).
 25. Howell, B. J., Hoffman, D. B., Fang, G., Murray, A. W. & Salmon, E. D. Visualization of Mad2 dynamics at kinetochores, along spindle fibres, and at spindle poles in living cells. *J. Cell Biol.* **150**, 1233–1250 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Bowerman, J. Sekelsky, N. Salmon, A. Harris and K. Bloom for comments on the manuscript; C. Waterman-Storer for discussions; the Cell Division Group at the Marine Biological Laboratory in Woods Hole; and B. Howell, D. Cimini, J. DeLuca, K. Shannon, C. Pearson, B. Moree and all members of the Salmon laboratory for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.D.S. (tsalmon@email.unc.edu).

S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex

Yuki Katou^{1,2}, Yutaka Kanoh^{1,2}, Masashige Bando¹, Hideki Noguchi¹, Hirokazu Tanaka^{1,3}, Toshihiko Ashikari⁴ & Katsunori Sugimoto⁵
Katsuhiko Shirahige¹

¹Genome Structure and Function Team, Human Genome Research Group, RIKEN Genomic Science Center, 1-7-22 Suehiro-cho, and ²Science of Biological Supramolecular Systems, Graduate School of Integrated Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

³Division of Biochemistry, Kihara Institute for Biological Research and Graduate School of Integrated Science, Yokohama City University, 641-12, Maioka-cho, Totsuka-ku, Yokohama, Kanagawa, 244-0813, Japan

⁴Institute for Advanced Technology, Suntory Limited, 1-1-1, Wakayamadai, Shimamoto-cho, Mishima-gun, Osaka, 618-8503, Japan

⁵Division of Biological Science, Nagoya University Graduate School of Science, Nagoya, 464-0814, Japan

The checkpoint regulatory mechanism has an important role in maintaining the integrity of the genome^{1–5}. This is particularly important in S phase of the cell cycle, when genomic DNA is most susceptible to various environmental hazards^{3,6,7}. When chemical agents damage DNA, activation of checkpoint signalling pathways results in a temporary cessation of DNA replication. A replication-pausing complex is believed to be created at the arrested forks to activate further checkpoint cascades, leading to repair of the damaged DNA. Thus, checkpoint factors are thought to act not only to arrest replication but also to maintain a stable replication complex at replication forks^{6–9}. However,

the molecular mechanism coupling checkpoint regulation and replication arrest is unknown. Here we demonstrate that the checkpoint regulatory proteins Tof1 and Mrc1 interact directly with the DNA replication machinery in *Saccharomyces cerevisiae*. When hydroxyurea blocks chromosomal replication, this assembly forms a stable pausing structure that serves to anchor subsequent DNA repair events.

We developed a high-density oligonucleotide array that was able to analyse yeast chromosome VI at a 300-base pair (bp) resolution, in order to understand the interaction of checkpoint factors with the DNA replication machinery. A total of 13,982 oligonucleotides (25-base nucleotides) were designed to cover the entire length (270 kilobases (kb)) of the chromosome. In a log-phase culture we demonstrated the detection of Orc1, a subunit of the origin recognition complex, bound to every known origin on chromosome VI (refs 10, 11; Fig. 1a). This shows that our chromatin immunoprecipitation (ChIP)-chip method^{12,13} can detect protein-binding sites at a resolution of at least 600 bp. In the following analyses, we used cells that were arrested at the G1 phase by α -factor, and then released into S phase in the presence of 200 mM hydroxyurea (HU) to activate replication checkpoint. Under these conditions we analysed the location of Cdc45, a protein involved in initiation and elongation steps of chromosomal DNA replication, and Pol1, a subunit of DNA polymerase α (Fig. 1b). Both showed localization only to early replicating regions, near autonomously replicating sequences (ARS) 607, 606, 605 and 603.5, suggesting that our ChIP-chip method can monitor locations of two proteins simultaneously. To analyse dynamic aspects of replication-related proteins during replication, we determined replicated regions by 5-bromodeoxyuridine (BrdU) incorporation simultaneously with protein binding. We used a yeast strain engineered to take up BrdU efficiently¹⁴ to detect BrdU incorporation on the chip simultaneously with the location of Cdc45 (Fig. 1c). Locations of BrdU incorporation coincided clearly with Cdc45 binding. The replicated regions determined by BrdU incorporation in the presence of 200 mM HU were consistent with data obtained by quantitative polymerase chain reaction using a wild-type strain (data not shown).

Next, we determined the location of components involved in the S-phase checkpoint signal cascade (Ddc1, Mec3, Ddc2, Tof1, Mrc1, Rad9, Rad24 and Rad53 (refs 1, 2)) through challenge with 200 mM HU at S phase (Fig. 1d; see also Supplementary Fig. S1). We could not detect any significant signals for Rad9, and association of Rad24 was detected mainly near telomeres. We could not detect the association of Rad53 under these conditions; however, when we used DNA (dimethyl adipimidate)—a protein–protein crosslinking reagent¹⁵—before formaldehyde fixation, Rad53 association with replicating regions became barely detectable (Supplementary Fig. S1). Locations of all other checkpoint proteins overlapped with BrdU incorporation, suggesting that either these proteins moved together with replication forks, or that they were recruited to stalled forks in search of specific DNA structures. To address this question, protein-binding patterns during early replication in the absence of HU was determined. Cells were released from G1 arrest at low temperature (16 °C) in the medium without HU to slow the DNA synthesis, and collected every 10 min to analyse locations of Tof1, Pol1 (Fig. 2a, b) Ddc1, Ddc2 and Mrc1 (Supplementary Fig. S2). Tof1, Mrc1 and Pol1 localized to early replication origins, and moved away from replication origins as time passed. Co-localization of Tof1 or Mrc1 with Pol1 was estimated by the correlation coefficient of their patterns of localization for each time point. They were 0.70 (40 min), 0.79 (50 min) and 0.78 (60 min) between Pol1 and Tof1 localization, and 0.43 (40 min), 0.70 (50 min) and 0.89 (60 min) between Pol1 and Mrc1 localization. This suggests that Tof1 and Mrc1 are located in replication forks as components of the replication machinery during normal S-phase progression. Not all of the components of the S-phase

checkpoint cascade associated physically with the replication forks. For example, we could not detect any significant signals of Ddc1 or Ddc2 at any time points (Supplementary Fig. S2). Biochemical confirmation for the association of Tof1 and Mrc1 with the replication machinery was obtained by demonstrating that these

proteins interact physically with Cdc45 *in vivo*, preferentially during S phase (Fig. 3a–c).

As Mrc1 and Tof1 are supposed to be specifically required for regulation of the DNA replication checkpoint, we analysed the progression of DNA replication in *tof1* and *mrc1* deletion mutants

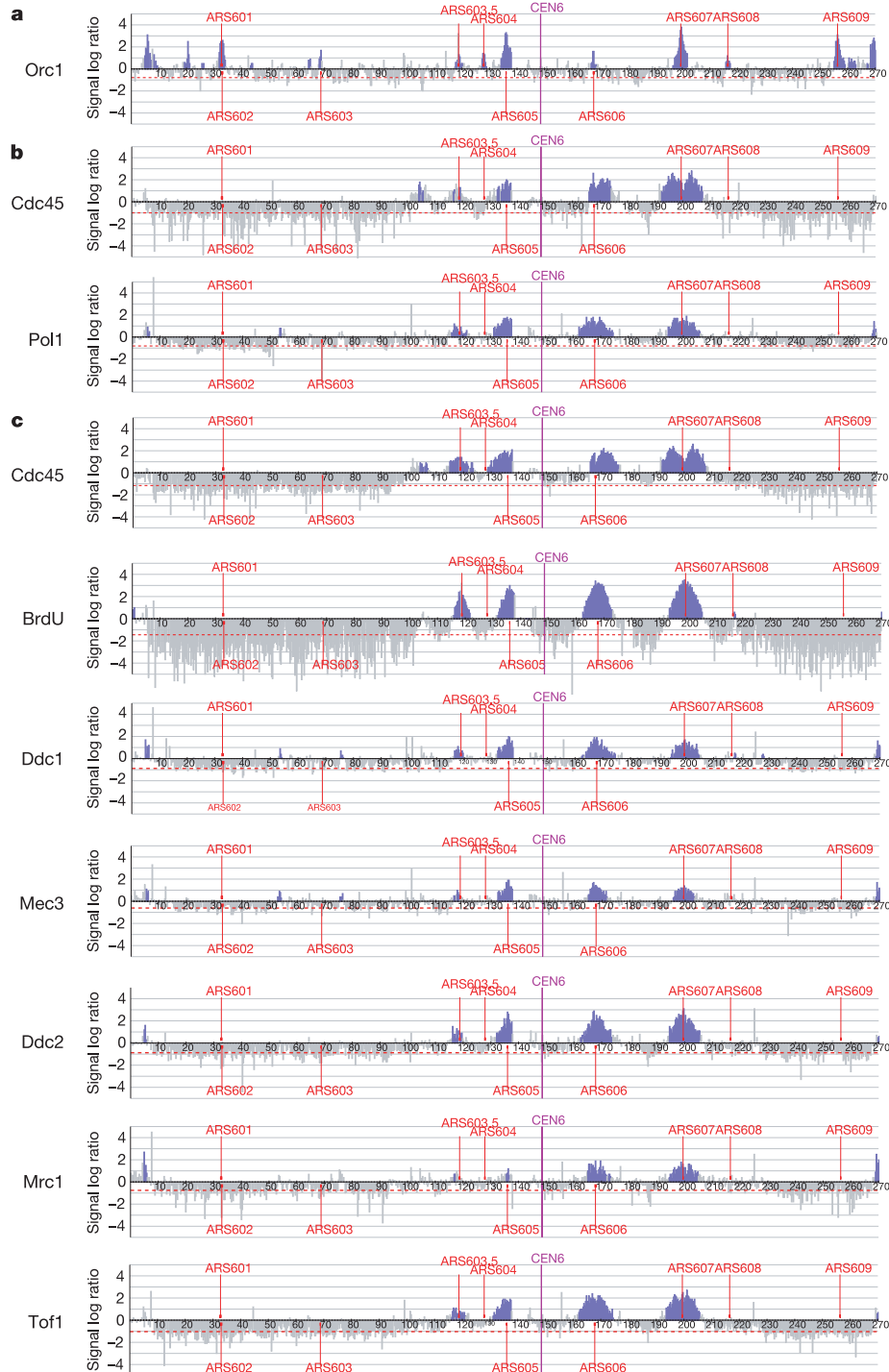


Figure 1 Locations of replication and checkpoint proteins on chromosome VI. **a**, Location of Orc1 protein on chromosome VI. Exponentially growing SKY002 (Orc1- $\times$ 3HA) cells were used for analysis. The blue shading represents the binding ratio of loci that show significant enrichment in the immunoprecipitated fraction. The red dotted line indicates the average signal ratio of loci that are not enriched in the immunoprecipitated fraction. The scale of the vertical axis is $\log_2$. The horizontal axis shows kilobase units (kb). **b**, Locations of Pol1 and Cdc45 in HU-arrested cells. SKY010

(Pol1- $\times$ 3Flag Cdc45- $\times$ 3HA) cells were used. **c**, Location of Cdc45 binding and BrdU-incorporated regions in cells arrested in early S phase by HU. Cells of SKY020 (Cdc45- $\times$ 3HA GPD-TK $_{7\lambda}$) were used as in Methods. **d**, Location of Ddc1, Ddc2, Mec3, Mrc1 and Tof1 protein binding in the presence of 200 mM HU. Regarding Tof1 and Mrc1, control ChIP-chip experiments were performed without prior crosslinking, and no locus was scored as significantly enriched.

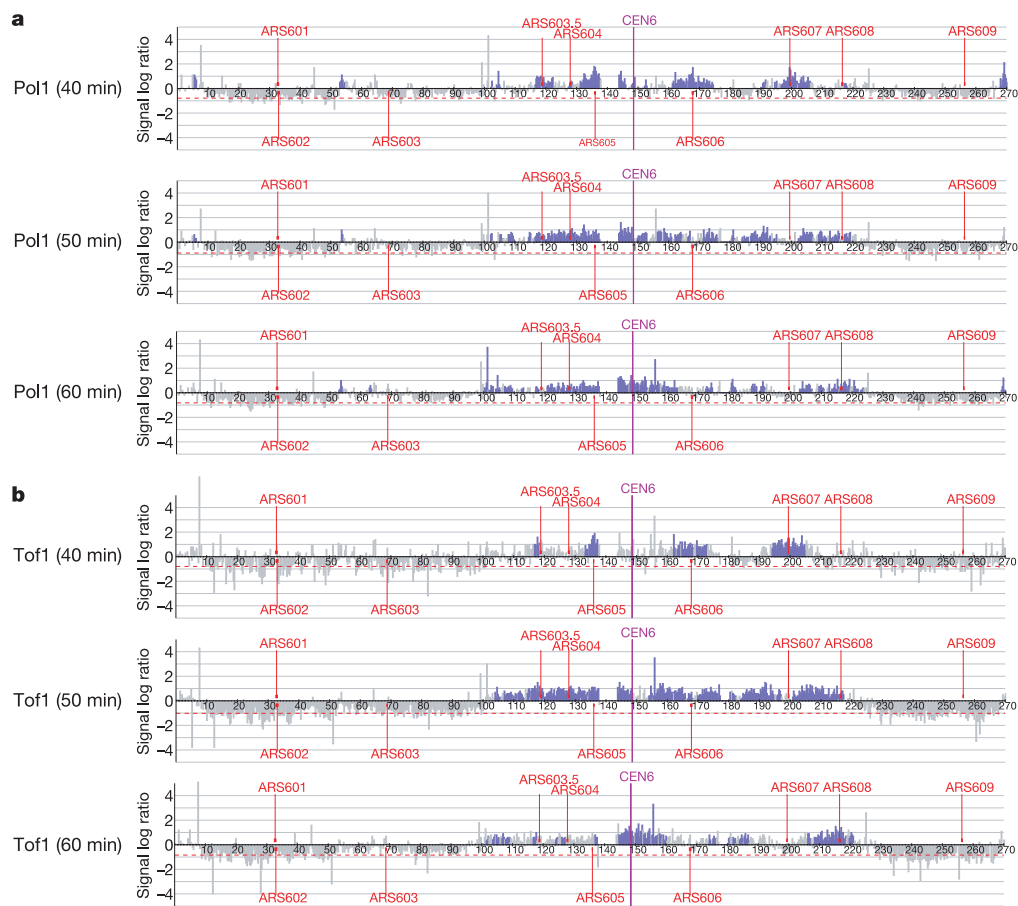


Figure 2 Interactions of Pol1 and Tof1 with chromosome VI during normal S-phase progression. **a**, **b**, Location of Pol1 (**a**; tagged by $\times 3$ Flag) and Tof1 (**b**; tagged by $\times 3$ Flag) after the release from α -factor arrest. Cells expressing a tagged version of each protein

were arrested in G1 by α -factor and then released into S phase by addition of pronase at 16 °C to slow down DNA synthesis. Cells were withdrawn every 10 min and the location of each protein analysed. The time after the release is indicated.

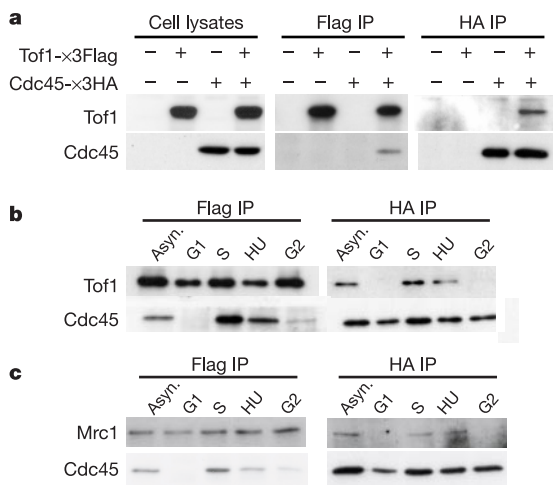


Figure 3 S-phase-specific association of Tof1 and Mrc1 with Cdc45. **a**, Co-immunoprecipitation of Tof1 and Cdc45. Left, extracts prepared from asynchronous cells with or without $\times 3$ Flag-tagged Tof1 or $\times 3$ HA-tagged Cdc45 were western blotted and probed with anti-Flag antibody (top) or anti-HA (bottom) antibody. Middle, same extracts as in the left panel were immunoprecipitated with anti-Flag antibody and probed with anti-Flag (top) or anti-HA (bottom) antibody. Right, same extracts as in left panel were immunoprecipitated with anti-HA antibody and probed with anti-Flag (top) or anti-HA (bottom) antibody. **b**, Co-immunoprecipitation of Tof1 and Cdc45 in S phase.

Immunoprecipitation with anti-Flag (Flag IP) or anti-HA (HA IP) antibodies of extracts from cells (Tof1- $\times 3$ Flag Cdc45- $\times 3$ HA) growing exponentially (Asyn.), synchronized in G1 by α -factor (G1), released from G1 into S phase (S), released from G1 into S phase in the presence of 200 mM HU (HU), or arrested in G2/M by nocodazole (G2). Tof1 and Cdc45 were then detected as in **a**. **c**, Co-immunoprecipitation of Mrc1 and Cdc45 in S phase. Immunoprecipitation with anti-Flag (Flag IP) or anti HA (HA IP) antibodies were performed on extracts from cells co-expressing $\times 3$ Flag-tagged Mrc1 and $\times 3$ HA-tagged Cdc45. Extracts were prepared and analysed as in **b**.

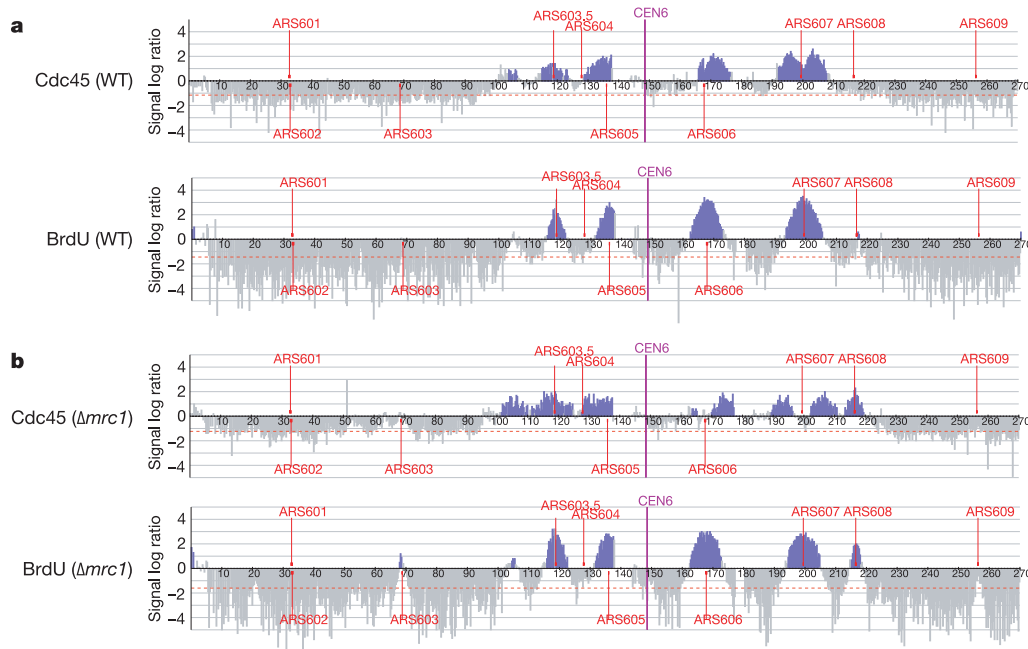


Figure 4 Location of Cdc45 and replicated regions in wild type (WT) and *mrc1* mutants treated with HU. **a, b**, Locations of Cdc45 are shown together with BrdU-incorporated regions in wild-type strain (**a**) and the *mrc1* deletion mutant (**b**). The mutant strain

originated from the wild-type strain SKY020 (Cdc45-x3HA GPD-TK_{7X}). Cell cultures were prepared as described in Methods.

by monitoring both BrdU incorporation and the location of Cdc45 in the presence of 200 mM HU. Figure 4 shows that in addition to replication from early origins, firing of late origins—ARS608 and ARS603—also occurred in a *mrc1* mutant as previously reported¹⁶, and also in a *tof1* mutant strain (Supplementary Fig. S3). The sizes of replicated regions from early origins were essentially the same as in wild-type cells. In contrast, the location of Cdc45 relative to replicated regions was different from that of wild type. We fitted the signal intensities to gaussian functions using the Levenberg–Marquardt method, and calculated the distance of protein-binding regions from replicated regions. In *mrc1* and *tof1* mutants, Cdc45 moved 2.5–3 kb further from ARS607 and ARS606 than in wild-type cells. In the *rad9* deletion mutant, replicated regions became narrower than in wild type; however, they were still closely coupled with locations of Cdc45 (Supplementary Fig. S3).

To study whether uncoupling of Cdc45 from DNA synthesis was due to the loss of Mrc1 and Tof1 function or resulted from defects in activation of the replication checkpoint, we analysed the replication profile in a *mec1 tel1* double-deletion mutant. We used this double-deletion mutant because *TEL1* is a functional homologue of *MEC1* (ref. 2). We used a *sm1* deletion mutant to allow survival of *mec1 tel1* deletion mutants¹⁷ (Supplementary Fig. S3). In the *sm1* mutant, replicated regions became larger, probably owing to elevated levels of dNTPs in the cell¹⁷. This did not affect the coupling of Cdc45 to DNA synthesis. In *sm1 mec1 tel1* triple mutants, sizes of replicated regions from early origins were almost the same as those of *sm1* deletion mutants. Cdc45 binding became hardly detectable around early origins, which could be the result of destabilization of replication forks in the *mec1 tel1* mutant strain^{6–9}. These results suggest that uncoupling of Cdc45 from DNA synthesis in *mrc1* and *tof1* mutants does not stem from a general failure to activate the checkpoint, but is a consequence of removing Mrc1 and Tof1 from the replication machinery.

To examine the uncoupling of Cdc45 from DNA synthesis, the time course of Cdc45 location was determined together with BrdU

incorporation in wild-type cells and in *mrc1* mutants in the presence of HU (Supplementary Fig. S4). Cdc45 is always located in front of replicated regions in wild-type cells, whereas the appearance of two separate peaks of Cdc45—indicative of uncoupling from replicated regions—occurred as early as 40 min in *mrc1* mutants (Supplementary Fig. S4). These results indicate that uncoupling of Cdc45 from DNA synthesis occurred as soon as DNA synthesis was arrested. Furthermore, all other replication proteins that we tested—Pol1, Pol3 (subunit of DNA polymerase δ), Dpb3 (subunit of DNA polymerase ϵ), Rfa1, Cdc47 (Mcm7 subunit of Mcm helicase) and Cdc54 (Mcm4)—were found to be uncoupled from DNA synthesis after 60 min of treatment with HU in both *mrc1* (Supplementary

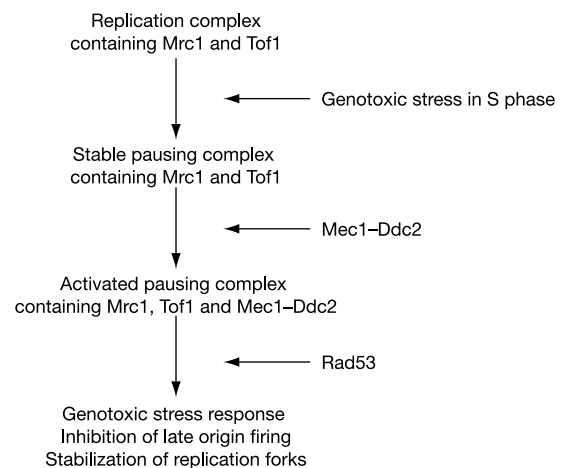


Figure 5 A model for the function of Mrc1 and Tof1. See text for details.

Fig. S5) and *tof1* mutants (data not shown). As it is known that Mrc1 proteins are not reloaded at replication forks once they have dissociated from the chromatin¹⁸, these data strongly suggest that the core replication machinery progresses away from the replicated regions in *mrc1* and *tof1* mutants, probably exposing non-replicated single-stranded DNA.

Rad9 becomes essential for S-phase checkpoint activation in the absence of Tof1 or Mrc1 (refs 16, 19). Consistent with this, we found that Rad9 is recruited to replicated regions in both *mrc1* (compare Supplementary Figs S1 and S5) and *tof1* mutants (data not shown), but not in wild-type cells. Binding of Rad9 to replicated regions in *mrc1* and *tof1* mutants strongly suggests that damage-like structures caused by the uncoupling of replication machinery from DNA synthesis could subsequently activate the DNA damage checkpoint. This uncoupling may not elicit the right type of checkpoint signal for the activation of Mec1–Ddc2 and Rad53, causing Rad9 to take over the role of checkpoint activation in this situation. Mrc1 and Tof1 were loaded onto DNA independently of each other (Supplementary Fig. S5), but each protein became uncoupled from DNA synthesis when the other was missing. A double-deletion mutant of *tof1* and *mrc1* showed a more severe growth defect and higher sensitivity to HU than either of the single mutants. These results suggest that Mrc1 and Tof1 are not in the same replication checkpoint cascade.

On the basis of these findings, we propose that Tof1 and Mrc1 proteins are components of the replication machinery and move as DNA synthesis proceeds during normal S phase (Fig. 5). When replication is arrested by environmental stress such as HU, they act to build a pausing complex by preventing the uncoupling of the replication machinery from DNA synthesis. The Mec1–Ddc2 complex is then recruited to this structure and may phosphorylate Mrc1 or Tof1 to activate this stable pausing complex¹⁶. The activated pausing complex then recruits and activates Rad53, which subsequently elicits stress responses^{16,20}. Rad9 has a role that is distinct from Mrc1 and Tof1, because its function is not required for proper pausing of replication forks, but rather for the recognition of damage-like structures generated when Mrc1 and Tof1 function is compromised. □

Methods

Yeast strains and plasmids

All strains and plasmids used in this study are listed in Supplementary Information (see also <http://everythingchromosomeVI.gsc.riken.go.jp>) and are available on request. All were derivatives of BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) (Invitrogen). All epitope tags were integrated at chromosomal loci. In all cases except for RAD9, X3-haemagglutinin (HA) or X3-Flag tag was fused to the carboxy terminus of the protein of interest. In the case of RAD9, we fused X2HA tag to its amino terminus by integrating a vector Ylp128-2HA-RAD9. A X6-His plus X3HA or X6His plus X3Flag tag was fused to the C terminus of the protein of interest using the cassette amplified from pU6H3HA (ref. 21) or pU6H3Flag. The plasmid for expression of herpes simplex virus thymidine kinase gene was a gift from E. Schwob. This GPD-TK construct was recloned in an integrative vector pAUR101 (Takara Shuzo).

Yeast cultures

For the preparation of HU-arrested cells, we synchronized yeast cells with α -factor (2 μ M) and then released cells into 200 mM HU for 60 min at 23 °C as previously described²². For the simultaneous analysis of BrdU incorporation, BrdU was added to the culture 30 min before release from G1 arrest to a final concentration of 200 μ g ml⁻¹, as previously described¹⁴.

ChIP on chip analysis

A chromosome VI high-density oligonucleotide chip was produced by Affymetrix Custom Express Service (rikDACE, P/N 510636, Affymetrix). Sequence and position of oligonucleotides on the chip are available from Affymetrix or from <http://everythingchromosomeVI.gsc.riken.go.jp>. The DNA chip itself can be purchased from Affymetrix. Sixteen 25-nucleotide probes were selected evenly from each 300-bp locus on the basis of their uniqueness, to avoid cross hybridization and to make the uniform hybridization condition. Chromatin immunoprecipitation was carried out as described previously with a few modifications¹². We disrupted 1.5×10^8 cells by Multi-beads shaker (MB400U, Yasui Kikai), which was able to keep cells precisely at lower than 6 °C during disruption by glass beads. This enabled us to retrieve routinely more than 60% of tagged proteins to soluble fraction. Anti-HA monoclonal antibody HA.11 (16B12) (CRP Inc.)

and anti-Flag monoclonal antibody M2 (Sigma-Aldrich) were used for chromatin immunoprecipitation. ChIP DNA was purified and amplified by random priming as described¹³. A total of 2 μ g of amplified DNA was digested with DNaseI to a mean size of 100 bp, purified, and the fragments were end-labelled with biotin-N6-ddATP²³. Hybridization, washing, staining and scanning were performed according to the manufacturer's instruction (Affymetrix). Primary data analyses were carried out using the Affymetrix Microarray Suite version 5.0 software to obtain hybridization intensity, fold change value, change *P*-value and detection *P*-value for each locus. Detailed information of the analyses is available at <http://www.affymetrix.com/support/>. For the discrimination of positive and negative signals for the binding, we compared ChIP fraction with supernatant fraction¹² by using three criteria. First, the reliability of strength of signal was judged by detection *P*-value of each locus ($P \leq 0.025$). Second, reliability of binding ratio was judged by change *P*-value ($P \leq 0.025$). Third, clusters consisting of at least three contiguous loci that filled the above two criteria were selected, because it was known that a single site of protein–DNA interaction will result in immunoprecipitation of DNA fragments that hybridized not only to the locus of the actual binding site but also to its neighbours¹⁵. Control experiments using 1 ng purified genomic DNA as a starting template showed that 99% of signals were reliably detected (detection *P*-value ≤ 0.025) and essentially the same.

For the analyses of BrdU incorporation, the same fraction of cells used for ChIP were fixed by ice-cold buffer containing 0.1% azide, and then total DNA from 3×10^8 cells was purified. DNA was sheared to 300 bp by sonication, denatured, and mixed with 2 μ g anti-BrdU monoclonal antibody (2B1D5F5H4E2; MBL) as previously described^{24,25}. Antibody-bound and unbound fractions were subsequently purified, amplified, labelled and hybridized to the DNA chip. More detailed data are available at the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) database and at our website (<http://everythingchromosomeVI.gsc.riken.go.jp>).

Co-immunoprecipitation analyses

Co-immunoprecipitation analyses of Tof1 and Mrc1 with Cdc45 were performed as previously described²⁶.

Received 4 May; accepted 4 July 2003; doi:10.1038/nature01900.

- Zhou, B. S. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433–439 (2000).
- Melo, J. & Toczyski, D. A unified view of the DNA-damage checkpoint. *Curr. Opin. Cell Biol.* **14**, 237–245 (2002).
- Shirahige, K. *et al.* Regulation of DNA-replication origins during cell-cycle progression. *Nature* **395**, 618–621 (1998).
- Kondo, T., Wakayama, T., Naiki, T., Matsumoto, K. & Sugimoto, K. Recruitment of Mec1 and Ddc1 checkpoint proteins to double-strand breaks through distinct mechanisms. *Science* **294**, 867–870 (2001).
- Melo, J., Cohen, J. & Toczyski, D. Two checkpoint complexes are independently recruited to sites of DNA damage in vivo. *Genes Dev.* **15**, 2809–2821 (2001).
- Lopes, M. *et al.* The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* **412**, 557–561 (2001).
- Tercero, J. A. & Diffley, J. F. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature* **412**, 553–557 (2001).
- Tercero, J. A., Longhese, M. P. & Diffley, J. F. A central role for DNA replication forks in checkpoint activation and response. *Mol. Cell* **11**, 1323–1336 (2003).
- Sogo, J. M., Lopes, M. & Foiani, M. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* **297**, 599–602 (2002).
- Shirahige, K., Iwasaki, T., Rashid, M. B., Ogasawara, N. & Yoshikawa, H. Location and characterization of autonomously replicating sequences from chromosome VI of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, 5043–5056 (1993).
- Yamashita, M. *et al.* The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* **2**, 655–665 (1997).
- Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
- Iyer, V. R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
- Lengronne, A., Pasero, P., Bensimon, A. & Schwob, E. Monitoring S phase progression globally and locally using BrdU incorporation in TK(+) yeast strains. *Nucleic Acids Res.* **29**, 1433–1442 (2001).
- Kurdستاني, S. K., Roby, D., Tavazoie, S. & Grunstein, M. Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nature Genet.* **31**, 248–254 (2002).
- Alcasabas, A. A. *et al.* Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nature Cell Biol.* **3**, 958–965 (2001).
- Zhao, X., Muller, E. G. & Rothstein, R. A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools. *Mol. Cell* **2**, 329–340 (1998).
- Labib, K., Tercero, J. A. & Diffley, J. F. Uninterrupted MCM2–7 function required for DNA replication fork progression. *Science* **288**, 1643–1647 (2000).
- Foss, E. J. Tof1p regulates DNA damage responses during S phase in *Saccharomyces cerevisiae*. *Genetics* **157**, 567–577 (2001).
- Tanaka, K. & Russell, P. Mrc1 channels the DNA replication arrest signal to checkpoint kinase Cds1. *Nature Cell Biol.* **3**, 966–972 (2001).
- De Antoni, A. & Gallwitz, D. A novel multi-purpose cassette for repeated integrative epitope tagging of genes in *Saccharomyces cerevisiae*. *Gene* **246**, 179–185 (2000).
- Paulovich, A. G. & Hartwell, L. H. A checkpoint regulates the rate of progression through S phase in *S. cerevisiae* in response to DNA damage. *Cell* **82**, 841–847 (1995).
- Wenzler, E. A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
- Cimbora, D. M. *et al.* Long-distance control of origin choice and replication timing in the human β -globin locus are independent of the locus control region. *Mol. Cell. Biol.* **20**, 5581–5591 (2000).

25. Schubeler, D. *et al.* Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. *Nature Genet.* **32**, 438–442 (2002).
26. Zou, L. & Stillman, B. Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbp4p kinase. *Mol. Cell. Biol.* **20**, 3086–3096 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. De Antoni, E. Schwob, S. Bell and J. Hegemann for the plasmids; K. Tamai for the special preparation of BrdU antibody; H. Araki, K. Okumura, E. Winzler, M. Foiani, T. Yada and K. Umezū for their comments; E. Schwob, F. Uhlmann, R. Cha and H. Yoshikawa for critical reading of the manuscript; T. Itoh, Y. Nakao, A. Nakada and C. Kawagoe for technical assistance; and Y. Sakaki and all the members of the Genome Structure and Function team for their support. This work was supported partly by grants-in-aid on priority areas from the Ministry of Education, Culture, Sports, Science and Technology, Japan, to K.S. Y. Katou is a Junior Research Associate of the RIKEN GSC.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S. (shirahi@gsc.riken.go.jp). Data presented in this paper can be obtained from GEO (<http://www.ncbi.nlm.nih.gov/geo/>) with accession number GSE486.

Replication of a *cis-syn* thymine dimer at atomic resolution

Hong Ling¹, François Boudsocq^{2*}, Brian S. Plosky², Roger Woodgate² & Wei Yang¹

¹Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, and ²Laboratory of Genomic Integrity, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

* Present address: Institut de Pharmacologie et de Biologie Structurale, CNRS, 31077 Toulouse, France

Ultraviolet light damages DNA by catalysing covalent bond formation between adjacent pyrimidines, generating *cis-syn* cyclobutane pyrimidine dimers (CPDs) as the most common lesion¹. CPDs block DNA replication by high-fidelity DNA polymerases, but they can be efficiently bypassed by the Y-family DNA polymerase pol η ^{2,3}. Mutations in *POLH* encoding pol η are implicated in nearly 20% of xeroderma pigmentosum, a human disease characterized by extreme sensitivity to sunlight and predisposition to skin cancer^{4–6}. Here we have determined two crystal structures of Dpo4, an archaeal pol η homologue, complexed with CPD-containing DNA, where the 3' and 5' thymine of the CPD separately serves as a templating base. The 3' thymine of the CPD forms a Watson–Crick base pair with the incoming dideoxyATP, but the 5' thymine forms a Hoogsteen base pair with the dideoxyATP in *syn* conformation. Dpo4 retains a similar tertiary structure, but each unusual DNA structure is individually fitted into the active site for catalysis. A model of the pol η –CPD complex built from the crystal structures of *Saccharomyces cerevisiae* apo-pol η and the Dpo4–CPD complex suggests unique features that allow pol η to efficiently bypass CPDs.

From bacteria to humans, most ultraviolet-induced photoproducts are removed by well-characterized DNA repair mechanisms⁷. Photoproducts remaining during DNA replication are either avoided by recombination-dependent template switching or bypassed by specialized low-fidelity polymerases that can synthesize DNA directly opposite lesions⁸. Crystal structures of three Y-family polymerases^{9,10}, including a large fragment of *S. cerevisiae* pol η and two archaeal DinB homologues, Dbh and Dpo4 from *Sulfolobus solfataricus*, revealed a right-hand-like catalytic core composed of palm, thumb and finger domains as found in all

DNA polymerases, and a unique little finger domain at the carboxy terminus. In the ternary complex of Dpo4, primer-template and an incoming nucleotide, the thumb and little finger domains hold the DNA duplex, and the active site formed between the palm and finger domains is spacious and solvent-accessible owing to the unusually small thumb and finger domains¹¹.

Dpo4 exhibits robust polymerase activity on undamaged DNA and can by-pass a variety of DNA lesions, including CPDs¹². By adjusting the register of the 3' end of the primer strand opposite a CPD-containing template and using dideoxyATP (ddATP) to reduce nucleotidyl transfer, we have crystallized Dpo4 with the 3' (TT-1) or 5' (TT-2) thymine of the CPD base-paired with ddATP and poised for DNA synthesis (Fig. 1). These co-crystal structures of Dpo4 and CPD-containing DNA, determined at 2.3 Å and 2.0 Å resolution (Table 1), provide the first glimpse of how a DNA polymerase replicates severely distorted DNA.

The Dpo4 structures in TT-1 and TT-2 are nearly identical except for a less than 1 Å shift of the little finger domain, which is probably due to differences in the local structure of the DNA as well as different crystal lattice contacts (Fig. 1 and Table 1). Compared with the structure of Dpo4 complexed with undamaged DNA in the same space group as TT-1 (ref. 11), the thumb and little finger domains in both TT-1 and TT-2 move by 1–2 Å leading to a concomitant shift of the DNA duplex (Supplementary section 1). Notably, this movement of protein and DNA appears to pivot around the catalytic carboxylate triad and results in a similar arrangement of DNA substrate in the active site regardless of the nature of the templating base (Fig. 1e). The three catalytic carboxylates—Asp 7, Asp 105 and Glu 106—and the two metal ions are superimposable in TT-1 and TT-2 (Fig. 1c). Although the DNA duplex and the ddATP are shifted, the 3' OH group of the primer strand remains within 4 Å of the α -phosphate of the incoming nucleotide.

A covalently linked thymine dimer imposes severe steric hindrance on DNA polymerases. For the A- and B-family DNA polymerases, only one templating base is accommodated in the active site for each cycle of primer extension and the adjacent 5' base is flipped out¹³. Covalently linked CPDs cannot separate and therefore do not fit into the 'closed-off' active site of replicative polymerases. In contrast, when the 3' base of the thymine dimer serves as a template in TT-1, the entire CPD slips into the open active site of Dpo4 (Fig. 1c, d). The 3' thymine and the incoming ddATP, although not perfectly co-planar, form two Watson–Crick-type hydrogen bonds of 2.6–2.7 Å in length (Fig. 2). The 5' thymine of the CPD is only 2.7 Å away from the carbonyl oxygen of Gly 58. This glycine residue is replaced by Ser or Met in human and *S. cerevisiae* pol η . Although the Met or Ser side chain points towards the major groove and neither would clash with DNA (Fig. 3), Gly is

Table 1 Summary of crystallographic data

Crystal (space group)	TT-1 (P2 ₁ 2 ₁ 2)	TT-2 (P2 ₁ 2 ₁ 2 ₁)
Unit cell (a, b, c) (Å)	98.5, 101.6, 52.2	98.7, 102.3, 106.1
Number of complexes in AU	1	2
Non-hydrogen atoms	3,569	7,494
Resolution range (Å)†	19.34–2.28 (2.32–2.28)	19.9–2.00 (2.06–2.03)
R _{merge} ‡	0.070 (0.554)	0.044 (0.492)
Unique reflections	24,302	69,785
Completeness (%)†	98.9 (99.0)	99.4 (95.0)
R _{value} §	0.252	0.222
R _{free} §	0.282 (684 reflections)	0.256 (1,012 reflections)
r.m.s. deviation bond length (Å)	0.009	0.010
r.m.s. deviation bond angle (°)	1.72	1.70
Average B-value (Wilson; Å ²)	61.6 (66.5)	46.2 (43.1)

r.m.s., root mean square. AU, asymmetric units.

†R_{merge} = $\sum_i \sum_h |I_{hi} - \langle I_h \rangle| / \sum_h I_h$, where I_{hi} is the intensity of the i th observation of reflection h , and $\langle I_h \rangle$ is the average intensity of redundant measurements of the h reflections.

‡Data of the highest resolution shell are shown in parentheses.

§R_{value} = $\sum_i ||F_o| - |F_c|| / \sum_i |F_o|$, where F_o and F_c are the observed and calculated structure-factor amplitudes.

§R_{free} is monitored with the reflections in parentheses excluded from refinement.

less restricted in backbone conformation and thus more accommodating for the CPD. Indeed, replacing the equivalent Ser 62 in human pol η by Gly made the mutant polymerase bypass a CPD more efficiently¹⁴.

In TT-2, the 3' thymine of the CPD forms a canonical Watson–Crick base pair with the adenine at the 3' end of the primer strand, and the 5' thymine is opposite the incoming ddATP. The templating

thymine constrained by the inter-base covalent bonds, however, does not rotate 36° or rise by 3.4 Å relative to the preceding base pair as in normal DNA. If ddATP assumes its usual conformation for nucleotidyl transfer, it could not form a Watson–Crick base pair with the 5' thymine of the CPD. If it were to form the base pair, the triphosphate could not fit into its binding site for the chemical reaction. In TT-2, the incoming ddATP assumes a *syn* conformation

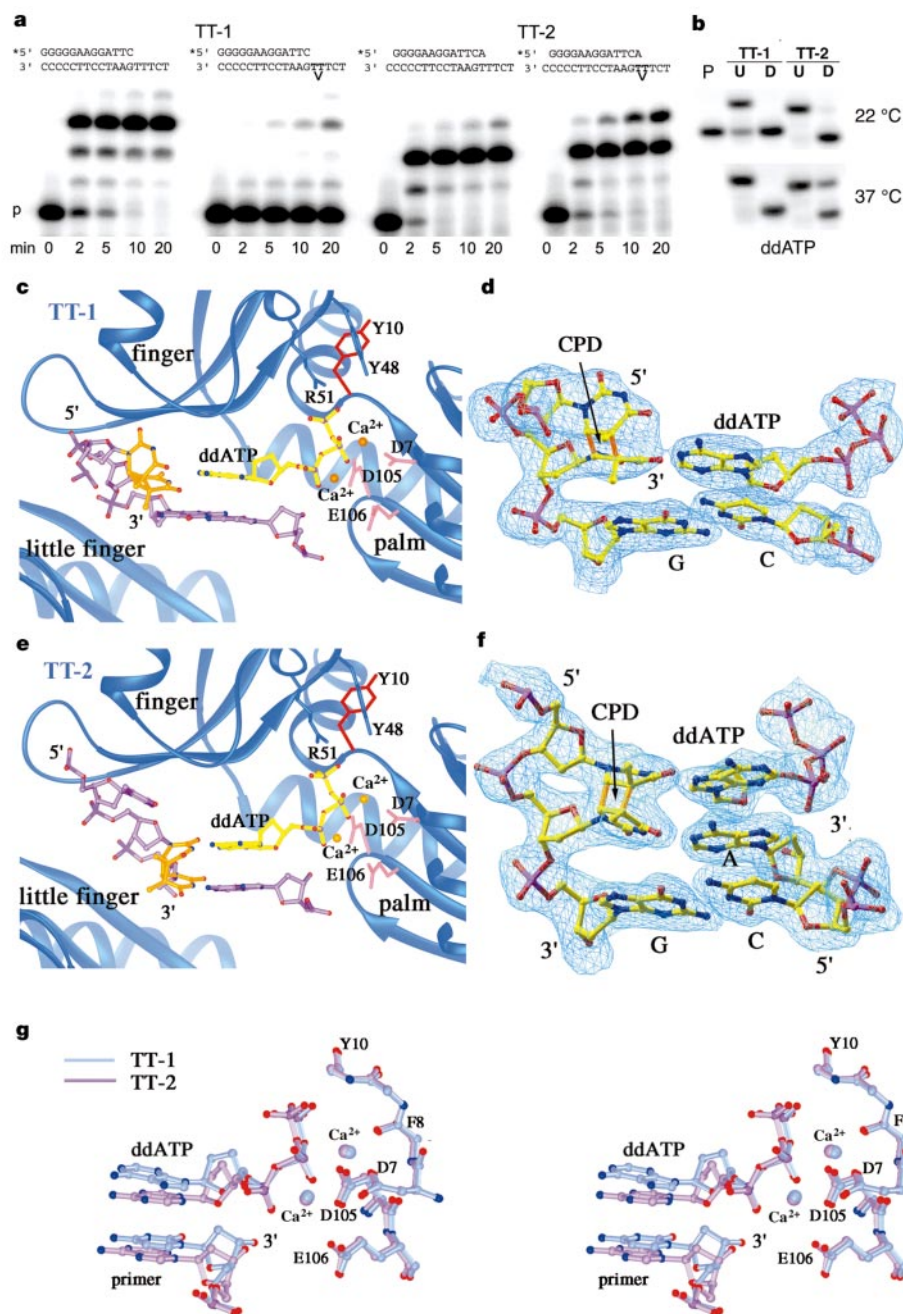


Figure 1 Replication of a CPD by Dpo4 in solution and crystals. **a**, Extension of two primers (13 nucleotides each) paired with undamaged or CPD-containing 18-nucleotide templates used in the crystallization studies (TT-1, TT-2). Reactions were carried out with 10 nM DNA substrate, 10 nM Dpo4 and 100 μ M dATP at 37 °C for 2, 5, 10 or 20 min. **b**, Inhibition of primer extension by ddATP. P indicates primer strand, and U and D indicate undamaged and CPD-containing template, respectively. The reactions took place for 30 min at 22 °C or 37 °C as indicated. **c**, The active site of TT-1, where the 3' thymine of the CPD (orange) is base-paired with ddATP (yellow). The conserved residues interacting with ddATP and catalytic carboxylates are highlighted. Tyr 10, which immobilizes the

finger domain by wedging between the palm and finger domains, is shown in red. **d**, Replication at the 3' T of the CPD. The CPD and the replicating and preceding base pairs of TT-1 are shown with the $F_o - F_c$ omit electron densities. **e**, Active site of TT-2, where the 5' thymine of the CPD is base-paired with ddATP, **f**, The replicating and two preceding base pairs of the TT-2 are shown with the $F_o - F_c$ omit electron densities. **g**, Stereo view of the TT-1 (blue) and TT-2 (pink) active-site superposition. The three catalytic carboxylates, two Ca^{2+} ions, the 3' nucleotide of the primer strand, and the incoming nucleotide are shown in the ball-and-stick model.

and forms a Hoogsteen base pair with the 5' thymine of the CPD, thereby maintaining the triphosphate contact with Dpo4 for nucleotidyl transfer (Fig. 1e, f and 2c). In the absence of the 3' OH of ddATP, the chemical bond does not form in the TT-2 crystal nor in solution at 22 °C (Fig. 1b). At 37 °C, incorporation of ddAMP opposite the 5' T of the CPD is much improved (Fig. 1b), and with dATP Dpo4 bypasses the 5' T of the CPD as efficiently as it does with undamaged DNA (Fig. 1a).

Hoogsteen base-pairing in duplex DNA has been observed when DNA is unwound or bound with intercalating drugs¹⁵, and occasionally when DNA is distorted by protein^{16–18}. In the crystal structure of a CPD-containing oligonucleotide¹⁹, Watson–Crick base pairs were retained, but the hydrogen bonds at the 5' T of the CPD were weakened and the DNA backbone of the strand opposite the CPD was distorted (Fig. 2a). To avoid backbone distortion, the Hoogsteen base pair at the 5' T of a CPD is probably favoured by all DNA polymerases for catalysis, which we believe is the first example of a Hoogsteen base pair facilitating an enzymatic reaction.

T to C and T to A mutations are observed at the 3' T of a CPD due to misincorporation of dGMP or dTMP, yet mutations rarely occur at the 5' T (refs 20, 21). When the 3' T of the CPD serves as a template, the close contact between the 5' T and the polymerase induces a 3 Å shift of the CPD towards the major groove, thus leaving ample space for the 3' T to form a wobble base pair with dGTP or dTTP, with each making two hydrogen bonds. At the 5' T of the CPD, however, the restraint imposed by the Hoogsteen base pair discriminates against guanine and pyrimidines, and only adenine, which makes two hydrogen bonds with thymine either by Watson–Crick or Hoogsteen base-pairing, is acceptable.

Dpo4 is rather inefficient at replicating the 3' T of a CPD (Fig. 1a),

whereas pol η replicates past CPDs as efficiently as undamaged DNA^{2,3}. To uncover the molecular basis for efficient bypass of CPDs by pol η , we generated a model of pol η complexed with a CPD based on TT-1 and the *S. cerevisiae* pol η structure²². With the TT-1 structure held fixed, pol η was brought in by superposition of its most conserved palm domain with that of Dpo4 (Fig. 3a). The finger, thumb and little finger domains of pol η , which bear many similarities to the corresponding domains of Dpo4, apparently need to rotate 16°, 16° and 45°, respectively, to be in position to interact with the DNA as the domains of Dpo4 do in TT-1 (Fig. 3a). In this modelled complex, pol η undergoes a 'closing' motion and makes numerous favourable contacts with the DNA. Even the Hoogsteen base pair fits well in the modelled pol η , when the TT-2 structure is superimposed onto TT-1.

Kinetic studies have suggested that pol η undergoes an induced-fit conformational change²³. Movement of the finger domain upon substrate binding is universal among the A- and B-family polymerases and is thought to be crucial for high-fidelity DNA replication¹³. Although conformational change of the finger domain in response to substrate binding has not been observed in the Dpo4 and Dbh structures^{10,24,25}, our model clearly indicates an induced fit of the finger, thumb and little finger domains in pol η . The finger domain is closest to the replicating base pair and participates in the active-site formation; therefore its movement is most significant.

Structure-based sequence alignment suggests that several amino acid substitutions in human and yeast pol η make their active sites much less exposed to solvent than that of Dpo4 (Fig. 3). The larger Gln and Ile side chains in place of Dpo4's Val 32 and Ala 44 potentially interact with the minor groove of the replicating base

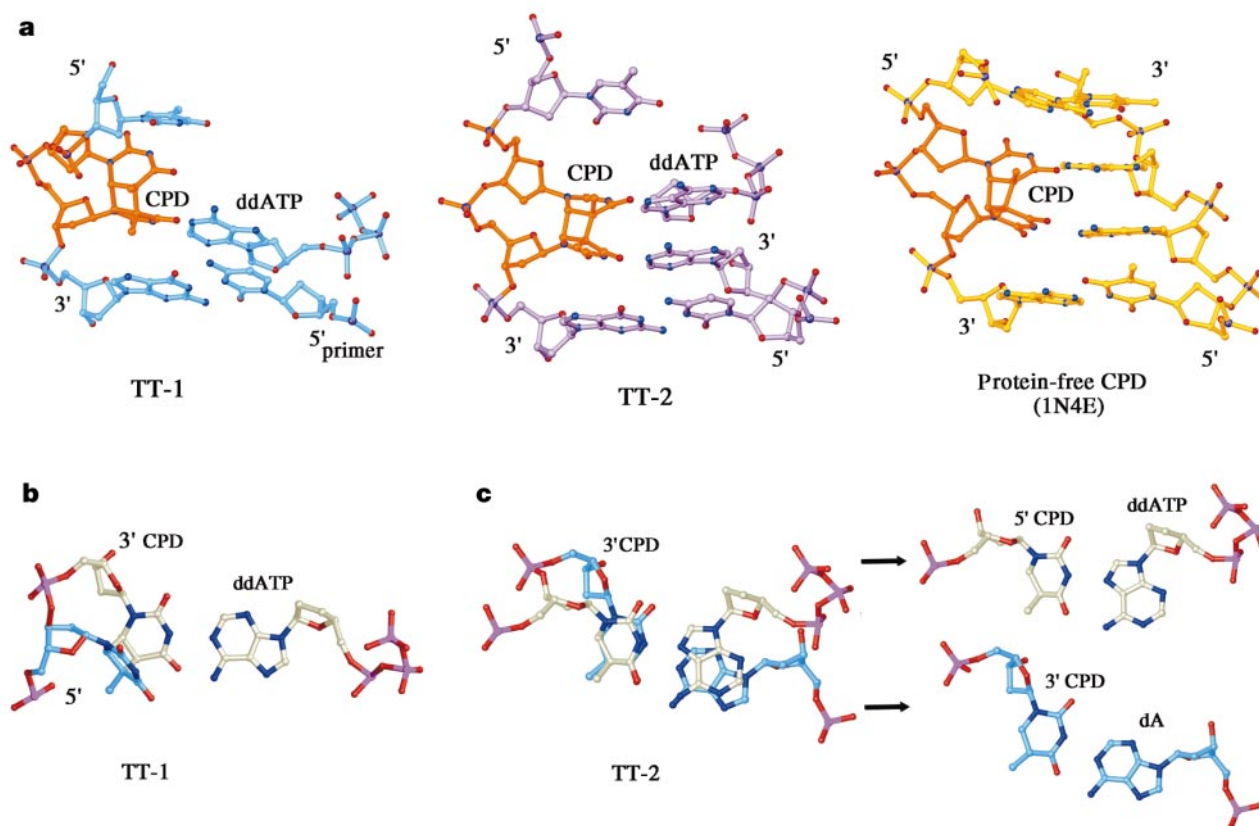


Figure 2 Structural comparison of the CPD complexed with Dpo4 and protein free. **a**, Structures of the CPD (orange) and surrounding nucleic acids in TT-1 (blue), TT-2 (pink) and in the absence of protein (yellow)¹⁹ are shown in the ball-and-stick model after superposition of the CPDs. **b**, **c**, Ball-and-stick presentations of base-pairing of the CPDs

in TT-1 (**b**) and TT-2 (**c**). The phosphorus atoms are shown in purple, oxygen in red and nitrogen in dark blue. The carbon atoms of the replicating base pair are white, and others are light blue.

pair and stabilize the CPD. The side chain of the Met or Ser that replaces Gly 58 of Dpo4 extends the protein–DNA interactions into the major groove. The Arg in place of Ala 57 of Dpo4 potentially orients the incoming nucleotide by charge interactions with the

β -phosphate, and the aliphatic portion of this Arg shields the major groove of the replicating base pair and the triphosphate moiety of the incoming nucleotide. Because of these amino acid substitutions in pol η , incoming nucleotides cannot enter the closed active site.

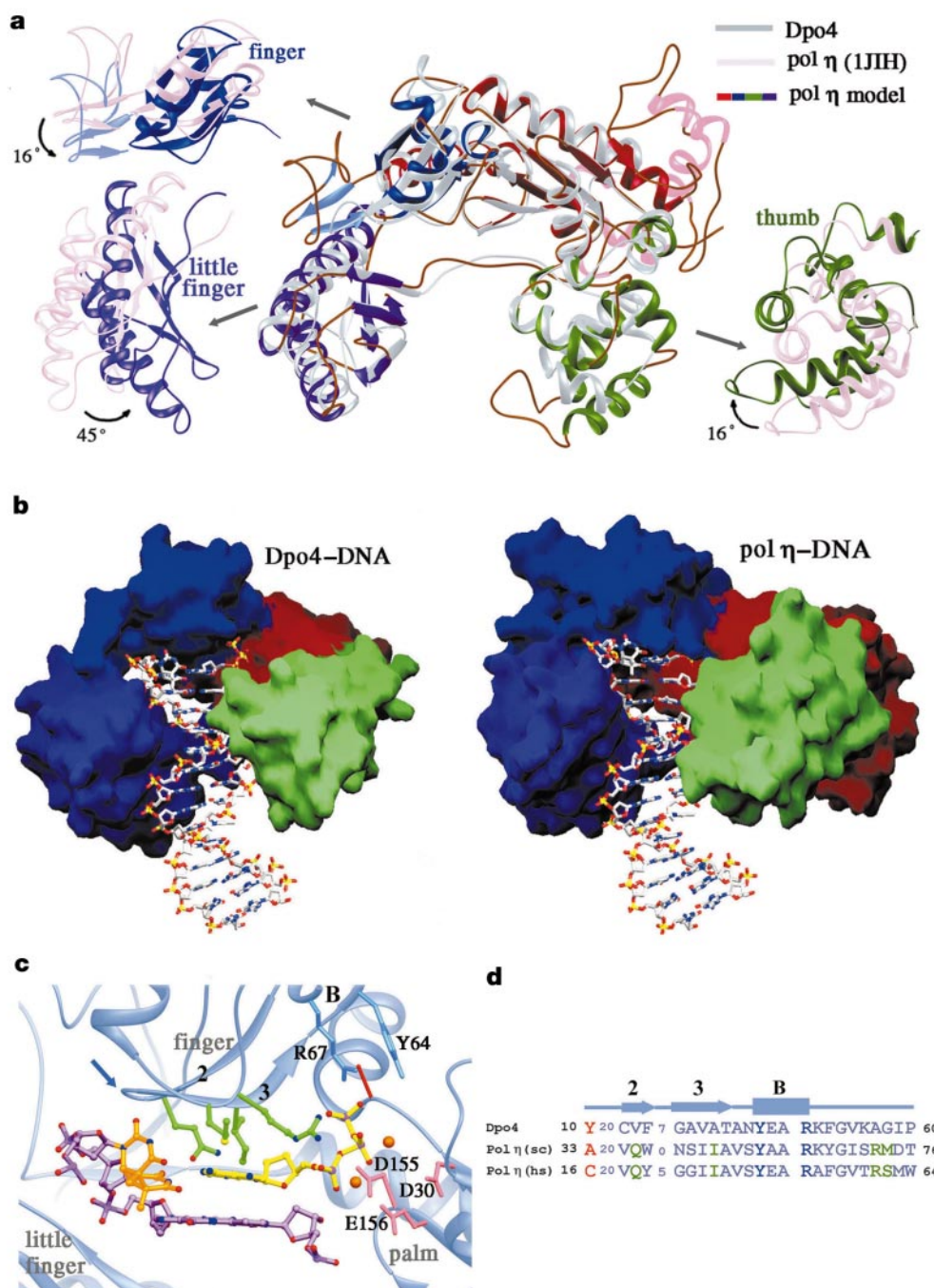


Figure 3 Modelling of *S. cerevisiae* pol η complexed with a CPD. **a**, Superposition of pol η (multicoloured) and Dpo4 (grey) after individual rigid-body movements of the pol η thumb, finger and little finger domains. Movements of the pol η domains between the apo-protein crystal structure (light pink) and the composite model are depicted in the surrounding panels. The pol η palm, thumb, finger and little finger domain are coloured red, green, blue and purple, respectively. The insertions unique in the pol η palm and finger domains are shown in pink and light blue. **b**, Comparison of the Dpo4 TT-1 crystal structure and the pol η –CPD model. The proteins are shown in molecular surface, and each domain is colour-coded as pol η in **a**. DNA and the incoming nucleotide are represented as stick models. The template and the incoming nucleotide in the pol η complex are less solvent-

accessible than those complexed with Dpo4. **c**, Close-up of the pol η active site. Pol η is shown as the light-blue ribbon diagram; the DNA substrate, the conserved protein residues and the two divalent cations are coloured as in Fig. 1c, e. Amino acid substitutions in human and *S. cerevisiae* pol η , which may enhance the efficiency for bypass of CPD, are highlighted in green. Ala 33 is highlighted in red. The blue arrow points out the deletion between β -strands 1 and 2 in pol η . **d**, Sequence alignment of the residues in finger domains of Dpo4, *S. cerevisiae* and human pol η that interact with the replicating base pair. The conserved residues are colour-coded as in **c**. The numbers in black are the beginning and ending residue numbers of each sequence; the numbers in blue indicate the intervening residues omitted in the alignment.

To open up the finger domain between successive rounds of polymerization, Tyr 10 of Dpo4, which immobilizes the finger domain by wedging between the palm and finger domains (Fig. 1), is replaced by Ala or Cys in pol η (Fig. 3d). The solvent-occluded active site together with the conformational change from the open DNA-free to closed DNA-bound form may, therefore, facilitate the ability of pol η to bypass CPDs. □

Methods

Preparation of protein and DNA

Dpo4 was prepared as reported previously¹¹. An 18-nucleotide DNA oligonucleotide (3'-CCCCCTTCTAAGTTTCT-5') containing a *cis-syn* thymine dimer (in bold) at the 14th and 15th positions from the 3' end was chemically synthesized and purified by Glen Research. Two complementary 13-nucleotide primer strands were purchased from Oligo Etc. Primer-1 (5'-GGGGAAGGATTC-3') base-paired with the first 13 bases of the CPD-containing template, and primer-2 (5'-GGGGAAGGATTCA-3') base-paired with the 2nd to 14th bases of the template including the 3' thymine of the CPD. The template and primer strands were annealed and mixed with purified Dpo4 at a molar ratio of 1.2:1. The final concentration of Dpo4 was 6 mg ml⁻¹ in a buffer of 20 mM HEPES (pH 7.0), 5 mM MgCl₂ and 150 mM NaCl.

Crystallization and structure determination

The TT-1 crystal with primer-1 was grown using the hanging-drop method with 15% PEG3350, 0.2 M calcium acetate, 2.5% glycerol at 20 °C after mixing the protein-DNA sample with 1 mM ddATP for 10 min at 22 °C before setting up the drop. The TT-2 crystal with primer-2 was grown similarly. Macroseeding was performed a couple of times to enlarge crystals. Crystals were flash-frozen in propane equilibrated with liquid nitrogen for storage and data collection. Diffraction data were collected using a Raxis IV mounted on a Rigaku RU200 generator and processed using HKL²⁶. The TT-1 crystal was isomorphous to the original Dpo4-DNA complex (space group P2₁2₁2), but the TT-2 crystal was in P2₁2₁2₁ space group and had two Dpo4-DNA complexes in each asymmetric unit (Table 1).

Both crystal structures were solved by molecular replacement²⁷ using the original Dpo4-DNA complex as a search model, and refined to 2.03 and 2.28 Å resolution, respectively^{27,28} (Table 1). The topology and parameter data of the CPD for refinement was generated from 1VAS²⁹. All protein residues are in the most favoured and allowed region of a Ramachandran plot.

Model building

Pol η was modelled onto the TT-1 structure by superimposing the palm domains of the two proteins³⁰. The other three domains were each adjusted as a rigid body to optimize their superposition with the Dpo4 structure. Superposition of the thumb domains was approximate owing to multiple insertions in yeast pol η . After adjusting a few side-chain torsion angles guided by the rotamer library, the resulting model of the pol η -DNA complex has no steric clashes, and pol η is transformed from an open DNA-free state to a closed DNA-bound conformation.

In vitro bypass of CPDs by Dpo4

The 5' [³²P]-labelled 13-nucleotide primers and unlabelled 18-nucleotide template, as described in the crystallization studies, as well as an undamaged 18-nucleotide template were annealed at a molar ratio of 1:2. Standard replication reactions of 10 µl contained 40 mM Tris-HCl at pH 8.0, 5 mM MgCl₂, 100 µM of ultrapure dATP or ddATP (Amersham Pharmacia Biotech), 10 mM DTT, 250 µg ml⁻¹ BSA, 2.5% glycerol, 10 nM 5' [³²P] primer-template DNA and 10 nM Dpo4. After incubation at 37 °C or 22 °C, reactions were terminated by the addition of 10 µl of 95% formamide/10 mM EDTA and the samples heated to 100 °C for 5 min. Reaction mixtures (5 µl) were subjected to 20% polyacrylamide/7 M urea gel electrophoresis and replication products visualized by PhosphorImager analysis.

Received 10 June; accepted 22 July 2003; doi:10.1038/nature01919.

Published online 6 August 2003.

1. Ravanat, J. L., Douki, T. & Cadet, J. Direct and indirect effects of UV radiation on DNA and its components. *J. Photochem. Photobiol. B* **63**, 88–102 (2001).
2. Masutani, C., Kusumoto, R., Iwai, S. & Hanaoka, F. Mechanisms of accurate translesion synthesis by human DNA polymerase η . *EMBO J.* **19**, 3100–3109 (2000).

3. Johnson, R. E., Washington, M. T., Prakash, S. & Prakash, L. Fidelity of human DNA polymerase η . *J. Biol. Chem.* **275**, 7447–7450 (2000).
4. Lehmann, A. R. Replication of damaged DNA in mammalian cells: new solutions to an old problem. *Mutat. Res.* **509**, 23–34 (2002).
5. Masutani, C. *et al.* The XPD (xeroderma pigmentosum variant) gene encodes human DNA polymerase η . *Nature* **399**, 700–704 (1999).
6. Johnson, R. E., Kondratieff, C. M., Prakash, S. & Prakash, L. hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science* **285**, 263–265 (1999).
7. Thoma, F. Light and dark in chromatin repair: repair of UV-induced DNA lesions by photolyase and nucleotide excision repair. *EMBO J.* **18**, 6585–6598 (1999).
8. Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
9. Ohmori, H. *et al.* The Y-family of DNA polymerases. *Mol. Cell* **8**, 7–8 (2001).
10. Yang, W. Damage repair DNA polymerases Y. *Curr. Opin. Struct. Biol.* **13**, 23–30 (2003).
11. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91–102 (2001).
12. Boudsocq, F., Iwai, S., Hanaoka, F. & Woodgate, R. *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DNA polymerase with properties akin to eukaryotic Pol η . *Nucleic Acids Res.* **29**, 4607–4616 (2001).
13. Patel, P. H., Suzuki, M., Adman, E., Shinkai, A. & Loeb, L. A. Prokaryotic DNA polymerase I: evolution, structure, and “base flipping” mechanism for nucleotide selection. *J. Mol. Biol.* **308**, 823–837 (2001).
14. Glick, E., Vigna, K. L. & Loeb, L. A. Mutations in human DNA polymerase η motif II alter bypass of DNA lesions. *EMBO J.* **20**, 7303–7312 (2001).
15. Park, J. Y. & Choi, B. S. NMR investigation of echinomycin binding to d(ACGTTAACGT)2: Hoogsteen versus Watson-Crick A.T base pairing between echinomycin binding sites. *J. Biochem. (Tokyo)* **118**, 989–995 (1995).
16. Aishima, J. *et al.* A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res.* **30**, 5244–5252 (2002).
17. Patikoglou, G. A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217–3230 (1999).
18. Bunting, K. A. *et al.* Crystal structure of the *Escherichia coli* dcm very-short-patch DNA repair endonuclease bound to its reaction product-site in a DNA superhelix. *Nucleic Acids Res.* **31**, 1633–1639 (2003).
19. Park, H. *et al.* Crystal structure of a DNA decamer containing a *cis-syn* thymine dimer. *Proc. Natl Acad. Sci. USA* **99**, 15965–15970 (2002).
20. Zhang, H. & Siede, W. UV-induced T → C transition at a TT photoproduct site is dependent on *Saccharomyces cerevisiae* polymerase η in vivo. *Nucleic Acids Res.* **30**, 1262–1267 (2002).
21. Szekeres, E. S. Jr, Woodgate, R. & Lawrence, C. W. Substitution of *mutAB* or *rumAB* for *umuDC* alters the relative frequencies of the two classes of mutations induced by a site-specific T-T cyclobutane dimer and the efficiency of translesion DNA synthesis. *J. Bacteriol.* **178**, 2559–2563 (1996).
22. Trincão, J. *et al.* Structure of the catalytic core of *S. cerevisiae* DNA polymerase η : implications for translesion DNA synthesis. *Mol. Cell* **8**, 417–426 (2001).
23. Washington, M. T., Prakash, L., Prakash, S. & Yeast, D. N. A. polymerase η utilizes an induced-fit mechanism of nucleotide incorporation. *Cell* **107**, 917–927 (2001).
24. Zhou, B. L., Pata, J. D. & Steitz, T. A. Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol. Cell* **8**, 427–437 (2001).
25. Silvin, L. F., Toth, E. A., Pham, P., Goodman, M. F. & Ellenberger, T. Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*. *Nature Struct. Biol.* **8**, 984–989 (2001).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
27. Brünger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Jones, T. A., Zou, J.-Y. & Cowan, S. W. Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
29. Vassilyev, D. G. *et al.* Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **83**, 773–782 (1995).
30. Cohen, G. E. ALIGN: a program to superimpose protein coordinates. *J. Appl. Crystallogr.* **30**, 1160–1161 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Craigie and D. Leahy for critical reading of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.Y. (Wei.Yang@nih.gov). Coordinates have been deposited in the Protein Data Bank under accession codes 1PM0 and 1PM8.

Electrophoresis has come a long way

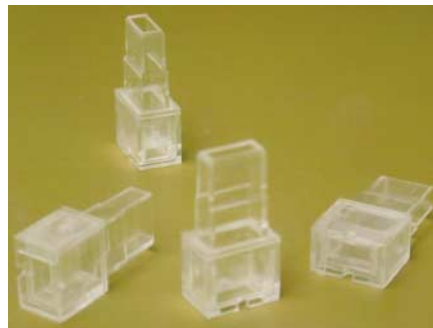
New ideas include a bespoke digital camera and calculating software.

GeneCAPSULE

Genotech www.genotech.com

One-step DNA extraction

GeneCAPSULE is a single-use device for the rapid recovery of DNA, RNA and protein from agarose or acrylamide gels. It uses electro-elution and requires no additional equipment. The device will extract DNA, RNA or protein as small as 20 bp or 4,000 molecular weight cut-off. The GEL-Pick is used to pick up the DNA, RNA or protein, which is then assembled using the GeneTrap. Once assembled, the GeneCAPSULE is submerged under the electrophoresis buffer and electro-elution occurs. The sample is recovered by piercing a small hole in the collection port and inserting a pipette tip for collection. Sample volume is in the range of 25–40 μ l and recovery rates reach 90%. The system takes as little as 60 seconds to recover 1,000 bp DNA. Samples are then ready for ligation, restriction enzyme digestion, sequencing, amplification, random priming, nick translation, enzymatic reactions and other applications.



GeneCAPSULE: just pop one of these on your gel.

CouloChem III

ESA Analytical www.esainc.com

Climate control for chromatography

ESA Analytical has added a temperature-controlled integrated organizer module to its CouloChem III HPLC electrochemical detector. The module is intended to save bench space and provide a protected, temperature-stabilized environment for columns and cells. The CouloChem III detector can operate in amperometric, coulometric, redox, screen, pulse or scan modes. DC and pulse/scan modes are in discrete modules to maximize performance. The detector can be used with any HPLC system for measurement of picogram to femtogram levels of oxidizable or reducible neurochemical, clinical, pharmaceutical, environmental or biological compounds.

Immun-Blot

Bio-Rad www.discover.bio-rad.com

Handy shortcut for western blotting

Immun-Blot pre-assembled filter paper sandwiches are designed to save time when performing western blotting and protein binding assays. They consist of a piece of polyvinylidene difluoride (PVDF) blotting membrane between two pieces of filter paper. The sandwiches are pre-cut to fit Bio-Rad precast gels as well as gels hand-cast by customers using the Bio-Rad mini-PROTEAN or Criterion cells. Sandwich sizes are 8.5 $\times$ 13.5 cm for Criterion gels and 7.0 $\times$ 8.5 cm for Ready Gels. Bio-Rad also offers pre-assembled nitrocellulose and Sequi-Blot PVDF blotting sandwiches.

pK_a Estimator, Log P Calculator, PRO Size

CombiSep www.combisep.com

Application-specific software

CombiSep has introduced three new software programs for electrophoresis applications. pK_a Estimator automatically extracts and displays 96-capillary electropherogram data for

user verification of labelled sample and marker peaks. pK_a is determined using nonlinear regression analysis. Log P Calculator can be used to determine compound octanol–water partition coefficients using multiplexed microemulsion chromatography. A typical electropherogram contains a sample peak bracketed by two migration markers. The software automatically extracts each electropherogram, establishes a system calibration curve and calculates log P values. Finally, PRO Size can be used to calculate molecular weights of protein samples based on the migration times of standards. Data are displayed in an overlay view, a stacked view, or as a simulated slab gel image. Reports can be exported containing the sample electropherogram, a simulated gel image and other pertinent sample information, including molecular weight and per cent purity.

BioDocAnalyze

Biometra www.biometra.de

Gels go digital

The BioDocAnalyze high-resolution (4 million pixels) digital camera for gel documentation can also be used for other tasks in the laboratory. All camera functions are controlled by the self-explanatory BioDocAnalyze software. A selection of pre-defined camera settings is provided for different gel types and stainings. In addition, individual user profiles can be defined. Other features include live preview, motor zoom and autofocus. BDAdigital is available with a small darkhood or with compact darkhoods for advanced applications.

PARIS

Ambion www.ambion.com

Capital idea

The new PARIS kit (a protein and RNA isolation system) uses a fast, simple, non-phenol-based procedure to isolate both RNA and native protein from the same experimental sample. The kit also permits separation of nuclear and cytoplasmic fractions prior to RNA and/or protein isolation. The resulting RNA samples are suitable for downstream applications such as array analysis, RT-PCR and northern analysis. The protein fractions can also be used directly for most common applications, including functional assays, immunoprecipitation, western blotting and two-dimensional gel electrophoresis.

These notes are compiled in the Nature office from information provided by the manufacturers.

Material intended for publication can be sent to newproducts@nature.com.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Natureevents:

Paul Constant (4954)

Production Manager:

 Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: r.asami@naturejpn.com

naturejobs

Experience counts

Old scientists don't really retire, they just find somewhere new to work — and Vida Vambutas is trying to help them. A former instructor at Hunter College in New York, Vambutas is building a database to help older scientists to find work, and to help companies and universities to tap into their knowledge. The seeds for this fledgling Retired Scientists Cooperative (www.retiredscientists.org) were planted a few years ago, when Vambutas was wondering what to do when she reached retirement age. During a subsequent fellowship at Columbia-Presbyterian Medical Center in New York, she met other colleagues who were pondering the same question. They now sit on the cooperative's board.

The idea has met with some interest from the corporate world. Vambutas initially wrote to companies about hiring retired scientists on a temporary basis. They wrote back asking for a list of available scientists, which she didn't have. Since then, she has built up a database of more than 200 retired scientists, but is seeking to get more diversity into the roster, which is weighted towards microbiology.

The construction of such a database should be a win-win situation. The scientists listed have spent a lifetime accumulating knowledge. Companies and universities could benefit from tapping into that knowledge, without much expense, as retired scientists are only seeking short-term work. Some of that knowledge — especially old techniques that have fallen out of favour — may still be relevant to today's workplace.

If successful, the model could be used by other associations and disciplines. Scientists, after all, are always looking for more problems to solve. And even when they have full-time jobs, many scientists are interested in extracurricular activities as independent contractors.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS AND STUDENTS

Blurring the line between
bench and bedside p1090

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

From bench to bedside

The training takes longer, but those willing to invest extra time in getting to grips with both basic and clinical research can reap the benefits, not least in job satisfaction, says Karen Kreeger.

After completing his PhD coursework at the University of California, Los Angeles, Neil Shah could readily recite the molecular biological underpinnings of chronic myelogenous leukaemia (CML). But it was only after following a haematology–oncology clinician on rounds and meeting a CML patient that he felt he understood the disease.

“It put a human face to what I understood at the molecular level,” says Shah of his literally bench-to-bedside training. Shah, now a clinical investigator at the Los Angeles campus, uses that mixed ‘translational’ background to study the mechanisms by which some leukaemic cells become resistant to some chemotherapy drugs, in an attempt to develop ways to override that resistance.

During the past few years, government agencies, universities and private foundations — such as the Leukemia and Lymphoma Society, which funds Shah — have been steadily pushing for more such translational research (see ‘Foundation funding’, below right). These programmes mean that the number of opportunities is increasing for scientists who are willing to undergo a longer training period in order to grasp both basic and clinical research.

Irwin Arias — assistant to the director of the Office of Intramural Research at the National Institutes of Health, and professor of physiology and medicine at Tufts University School of Medicine in Boston — says that the clinical setting is becoming a popular occupational route for PhDs, in addition to MDs doing basic-science postdocs, and notes that the demand for MD/PhDs has been high. Arias has long held an interest in translational medical education: he started a course at Tufts in the 1980s to train PhD students, fellows and faculty members in pathophysiology. He most recently created and directs a course for the National Cancer Institute to demystify medicine for PhDs working there. Other institutes are considering offering similar courses that are specific to their research areas.



Irwin Arias: demystifying medicine for PhDs.



Anna Keating: combining basic and clinical science offers several benefits.

MULTIPLE CHOICE

There are several possible paths to a translational career. Some take the bench-to-bedside route, starting out with a PhD in basic research, but then adding clinical experience from fellowships, coursework and collaboration with physicians to bring their ideas to patients. Others go from bedside to bench: ‘physician-scientists’ — MDs or MD/PhDs who spend their time primarily in the clinic — seek to work more in the lab. Finally, a few, like Shah, come at the issue from both directions while aiming for a specific research niche.

Hayley McDaid took the bench-to-bedside approach. Her interest in translational medical research

started with her PhD studies at Queen’s University in Belfast, UK, continued as a molecular-oncology instructor at the Albert Einstein College of Medicine in New York, and was facilitated five years ago with a Susan G. Komen Breast Cancer Foundation fellowship.

What helped McDaid along the way was her ability to cultivate research relationships with already-established clinicians. She contacted Sridhar Mani, a physician also at Einstein, to collaborate on a clinical study using the molecules she is investigating.

“This also helps with getting funded,” she notes. Her advice is to embark on good relationships with outstanding clinicians for the early phases of translational research. Then build on that by applying for other types of funding, such as the translational-research fellowship from the American Association for Cancer Research and drug firm Amgen that funds her current research.

Another group travels in the other direction, getting medical training first, then becoming more involved in basic research. Italian MD Laura Pasqualucci, now assistant professor of clinical

Foundation funding

Some newly minted physicians choose careers in patient-based research, although their numbers have been declining since the mid-1980s.

But some foundations still fund this type of career path, such as the Damon Runyon–Lilly Clinical Investigator Award.

Similarly, the Wellcome Trust offers fellowships targeted at clinicians such as MDs and doctors of dental surgery to train in research that is either bench-oriented or that involves working with human subjects. But it doesn’t offer fellowships necessarily targeted at translational research.

As an MD/PhD, David Feldman of Ohio

State University says that his physician–scientist fellowship from the Howard Hughes Medical Institute (HHMI) is what gave him credibility as a basic scientist. This award helped him pull together what he was learning about molecular and cell biology in the postdoc with his background in physiology and pharmacology and his clinical experience.

The HHMI has since discontinued the physician–scientist programme, but only in order to fund scientists who are slightly senior to Feldman’s cohort of postdocs. Its new recipients are clinicians who already have well-established basic-science laboratories.

K.K.



David Feldman feels that the problem with translational science is that everyone has a different definition of it.

pathology at Columbia University and a Leukemia and Lymphoma Society special fellow, took that path when she moved from Italy to New York six years ago, to improve her molecular-biological skills.

Working in the lab of Riccardo Dalla-Favera, she studies genetic alterations specifically associated with B-cell diffuse large-cell lymphoma and how this might be used to identify new targets for diagnosis and therapy. She notes that, in her experience in Europe, it is easier to get funding for translational medical research than for basic research.

LOST IN TRANSLATION

Translational medicine might seem to be rewarding and full of opportunities, but there are a few issues to grapple with for those who span both areas. “The problem with translational science is that everyone has a different definition of it,” says David Feldman, an MD/PhD and director of heart failure and cardiac transplantation at Ohio State University in Columbus. This sometimes vague idea of what constitutes translational medical research puts the onus on the grant writer to make sure that both molecular-biologist and clinician reviewers understand the fundamental and applied sides of the proposed project.

Getting funding for translational research can be especially challenging for investigators early in their careers, says Jim Olson, an MD/PhD who is an assistant member at the Fred Hutchinson Cancer Research Center and assistant professor of paediatrics at the University of Washington, Seattle. “If you haven’t already translated a drug, reviewers are uncertain that you can do it,” he says. Olson is in the first year of his Damon Runyon–Lilly Clinical Investigator Award, seeking new treatments for brain tumours in children.

Many universities are trying to recruit translational investigators, but often their work doesn’t fit into the traditional tenure-track model, he says. For example, published studies in the translational area tend to

appear under collaborative-group authorship, whereas first-author papers are prized on the tenure track.

Anna Keating, an MD who is a fellow in paediatric haematology and oncology at the Memorial Sloan-Kettering Cancer Center in New York, has not seen collaborative authorship and tenure review as a problem for herself or her colleagues. She is now in the first year of her Leukemia and Lymphoma Society fellowship, testing strategies to enhance the cytolytic activity of donor T cells. She says that translational researchers can benefit from the mix of journals in which they could publish, as their papers can have a basic as well as a clinical slant: basic-science papers can be published first to get the first-author papers out in time for review. “But,” she says, “it could also be institution-dependent in whether this is a problem” — depending on whether the departmental faculty members are basic scientists or mainly clinicians.

Despite possible problems in finding a comfortable academic fit for translational research programmes, the field is coming into its own, with bedside observations informing benchtop discoveries and vice versa. This has been the case for Albert Koong, an assistant professor of radiation oncology at Stanford University and a Damon Runyon–Lilly Clinical Investigator Award recipient. He studies pancreatic tumours, which are deficient in oxygen — a condition called hypoxia — and is looking at the signalling pathways during hypoxia with a view towards developing therapies based on that. His field-crossing studies ask why there is such an early propensity for the pancreatic tumour to spread, and whether the molecular biology of hypoxia plays a role in early metastases. His work in the clinic and with clinicians steers the direction of his basic research and helps him to formulate good questions about the nature of pancreatic cancer.

The upshot, says Keating, is that translational research “needs bi-directional communication” — clinicians and basic researchers can help to solve each other’s experimental problems. ■

Karen Kreeger is a freelance science writer based in Media, Pennsylvania.

Web links

Leukemia and Lymphoma Society

♦ www.leukemia-lymphoma.org

Susan G. Komen Breast Cancer Foundation

♦ www.komen.org/grants

American Association for Cancer Research

♦ www.aacr.org/1605.asp

Damon Runyon–Lilly Clinical Investigator Award

♦ www.cancerresearchfund.org

Wellcome Trust

♦ www.wellcome.ac.uk